THE UNIVERSITY OF ALBERTA

THE CLASSIFICATION ABILITIES OF ELEMENTARY SCHOOL STUDENTS

by

© Charles Douglas Blackford

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF ELEMENTARY EDUCATION

EDMONTON, ALBERTA

FALL, 1970

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "The Classification Abilities of Elementary School Students," submitted by Charles Douglas Blackford, in partial fulfilment of the requirements for the degree of Master of Education.

ABSTRACT

The purpose of this study was to discover if any significant differences existed between elementary school pupils with respect to their ability to classify physical objects; such ability being measured in terms of number and type of classification criteria used and speed of performance of the classification exercise.

A sample of one hundred and eighty children, composed of thirty children from each grade (ninety each from the Edmonton Public and County of Strathcona school systems) was used in this study. Each student was assigned to a high, average, or low intelligence group within his grade level on the basis of I.Q. scores as recorded in his cumulative record or on the basis of the recommendation of his teacher.

The children in the sample were presented individually with a classification task based upon twelve different substances, each of which was contained within a square bottle. The children were encouraged to classify the contents of the bottles in as many ways as they could, and their performance was evaluated on the basis of speed and upon the total number of classifications made. A record was kept of the criteria used by each child in order to determine if there were any differences in criterion preference between grades, I.Q. groups, and sexes.

The data were subjected to analyses of variance. Decisions to accept or reject hypotheses were made at the

0.05 level of significance for F and "t" tests, and at the 0.10 level of significance for the Scheffé test.

The results of this analysis showed that there were no significant differences between boys and girls in classification abilities or criterion preference. With respect to intelligence, the high I.Q. group in general performed better than the average and low groups, especially on the power score variable, and did not prefer color as a criterion as much as the low I.Q. group. The largest differences in classification abilities were found between grade levels; the evidence indicating that a growth in such abilities occurs between grades one and three. This growth in classification ability then appears to taper off briefly but resumes from grades four to six. A difference in criterion preference is revealed with respect to grade level, with grades one and two preferring the criterion of color and grades three to six seeming to prefer the criterion of liquid vs. solid state.

The main implication of this study was that experiences designed to improve the classification abilities of children should be continued at least up to grade six since no peak in these abilities is reached before that time.

It is hoped that the classification task used in this study, or others similar to it, will be found useful by classroom teachers for determining the classification abilities of their students.

ACKNOWLEDGEMENTS

The writer welcomes the opportunity to express his sincere gratitude to the many people who helped to make this thesis a reality.

Special thanks are due to Dr. D. W. R. Wilson, chairman of the thesis committee, for his advice and constructive criticism throughout the past year.

Thanks go to Professor N. M. Purvis, Dr. T. Maguire, and to Dr. M. Horowitz for their interest in, and acceptance of this project.

The writer is also indebted to Dr. G. Cathcart for his invaluable assistance with the statistical analysis.

Recognition is given to the administrators, teachers, and students of the Edmonton Public and County of Strathcona school systems, without whose co-operation this study would have been impossible.

Special appreciation is given to my parents for their interest and support, and to my wife, Sharon, for her ever-present patience, encouragement, and assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION TO THE STUDY

THE PROBLEM

The province of Alberta introduced a new science curriculum into its elementary schools in September 1968. This "new" science curriculum differs in several fundamental ways from the "old" curriculum. One of the most important additions to the list of course objectives is the requirement that the development of students' inquiry skills be attempted. The statement of the new objectives concerning inquiry skills reads:

> As a result of science instruction, the elementary school pupil should:
> (a) develop the ability to inquire, i.e., ability to think and investigate science through the use of process skills (behaviors) such as observing, classifying, communicating and inferring (Curriculum Guide, 1969, p. 5). (See Appendix A for a complete list and description of the processes.)

This objective is not a minor one, but forms an integral part of the total course and is given a value equal to that of the acquisition of concepts. As stated in the Curriculum Guide (1969):

> The new elementary school science program has two fundamental but inseparable objectives. By emphasizing the development and use of inquiry skills as tools of investigation, the program is designed to enable the student to become an active and dynamic investigator of science. To have the student develop basic science concepts is a second aim (p. 5). . . . The intent is that the teacher will develop a balanced program where process

and content receive equal emphasis (p. 7).

Since the elementary science curriculum now has two fundamental objectives, it is imperative that both curriculum and student evaluation be interms of both these objectives. Traditionally, teachers have evaluated only the student's ability to recall memorized information. "If the new approach in science is to succeed, a radical change in evaluation procedures is essential (Curriculum Guide, 1969, p. 9)." This necessary, radical change in evaluation procedures, however, has become the fly in the ointment. The successful implementation of the inquiry approach to science depends upon a successful change in methods of evaluation. In the last ten years, in spite of all the attention given to the inquiry approach to science teaching, no one has yet developed and completely validated a test of science inquiry skills. Many educational organizations are attempting to solve this problem but, until they do, most of the burden of inquiry skill evaluation depends on the intuition of the classroom teacher.

It is suggested by the Curriculum Guide (1969, p. 9) that teachers make a checklist of inquiry skills for each student and, through individual observation and subjective judgment, rank the student according to his inquiry abilities. However, even with this method the teacher must use some criteria for making a subjective judgment. What are these criteria to be? What sort of behavioral skills related to each of the inquiry skills should the teacher expect to observe in grade one students, grade two students, and so on through to

grade six.

The answers to these questions can initially be found only by direct observation of the behavior of students working on tasks designed to require each of these process skills for their solution. In order to make a start on the determination of answers to these questions, the process skill of classification was chosen for examination and analysis in this study. Classification was chosen because, in the words of Rawson (1965), "The earliest of these abilities to develop, [i.e. the abilities involved in logical thought] appears to be simple classification (p. 44)." It is the intention of the writer to carry out further research on the remaining process skills in order, hopefully, to build a significant body of knowledge concerning all the process skills of inquiry. Such a body of knowledge may prove useful as an aid to the development of validated written tests of process skills for the elementary grades.

THE PURPOSE

The purpose of this study was three fold:

1. To discover and describe the present level of classification ability, in terms of the number of classifications made, for a sample of children in each I.Q. group in each of the grades one through six.

2. To discover if a relationship exists between the number of classifications made, grade level, and I.Q.

3. To discover some of the criteria normal children

use to classify objects, and to discover whether these criteria tend to change with grade level.

SIGNIFICANCE OF THE STUDY

In view of the generally inconclusive nature of the research on the classification abilities of children in grades one through six, and because of the necessity of evaluating students and science curricula partially on the basis of the development of process skills, it is hoped that this research will:

1. Give teachers a criterion by which they may judge the classification abilities of their students in grades one through six.
2. Indicate at which grade levels emphasis on classification skills can be most productive.
3. Indicate any relationships which may exist between classification ability and (a) sex, (b) I.Q.
4. Give some indication of the type of classification criteria children in different grades think of, and the order in which they think of them.
5. Prove useful in the validation of written tests of classification ability.

DEFINITIONS

The following five terms have restricted and special operational meanings for the purpose of this study, and shall be used as hereinafter defined:

1. Classification. The term classification is used as in the Curriculum Guide (1969): ". . . grouping articles, objects and ideas on the basis of some observable property or properties (p. 7)."

2. Intelligence. Intelligence is designated as high, average, or low according to the subject's most recent intelligence test. For the purpose of this study the highest 20 percent of the students in each grade will be designated the high I.Q. group. The lowest 20 percent of the students in each grade will constitute the low I.Q. group and the middle 60 percent of the students will make up the average I.Q. group.

3. Classification Power Score. Each "appropriate" (see Scoring, page 40, for details) classification criterion identified by the student will receive a value of ten. One point will be subtracted from this score for each mistake made in assigning a particular sample to a specific classification group. The classification power score will be the sum of these criterion values.

4. Classification Speed Score. The average time taken to think of each of the first five criteria and complete the sorting appropriate for such criteria will be the classification speed score (i.e. the sum of the time intervals for the first five classifications divided by five).

5. Composite Classification Score. The composite classification score will be obtained by dividing the power score by the speed score.

HYPOTHESES

1. There is no significant interaction between grade level and I.Q. with respect to:

   (a) Classification power score,

   (b) Classification speed score,

   (c) Composite classification score.

2. There is no significant interaction between grade level and sex with respect to:

   (a) Classification power score,

   (b) Classification speed score,

   (c) Composite classification score.

3. There is no significant difference between boys and girls in:

   (a) Classification power score,

   (b) Classification speed score,

   (c) Composite classification score,

   in each of the grades one through six.

4. There is no significant difference between low, average, and high I.Q. students in:

   (a) Classification power score,

   (b) Classification speed score,

   (c) Composite classification score,

   in each of the grades one through six.

5. There is no significant difference between students in:

(a) Classification power score,

(b) Classification speed score,

(c) Composite classification score,

in each of the grades one through six.

LIMITATIONS

Heisenberg has drawn to our attention what might be called the universal limitation of research in physics. As he stated, "We can no longer speak of the behavior of the particle independently of the process of observation (Heisenberg, 1958, p. 15)." It is worth considering that in a similar way, in education, the very process of observation creates a disturbance in the behavior one is attempting to measure. More specifically, however, when interpreting the findings of this particular study one must, in addition, keep in mind the following limitations:

1. Students from grades two to six have been involved in the "new" science curriculum for only two years. After the program has been in effect for six years, a study similar to this may produce different results for student performance on the classification task.

2. No screening of students for defects such as in color vision or form descrimination was carried out.

3. No special check on the equality of the grade and I.Q. groups, with regard to confounding variables such as socio-economic status or previous history, was made. It was assumed that the random selection of students ensured this equality.

4. Although the test of classification ability used has construct validity, no attempt was made to determine its reliability.

5. Not all the children in the study were tested with the same I.Q. test. For 18 percent of the children there was no I.Q. score available, so they were placed according to the judgment of their classroom teacher. Taking these facts into account, some discretion must be exercised in the interpretation of the results which involve grouping on the basis of I.Q.

6. Although an attempt was made to minimize the effect of the teacher variable it must not be discounted. For example, a difference of interest or training in science instruction may exist between primary and upper elementary teachers which may account for some of the differences in classification abilities found between students from these two areas.

Chapter 2

REVIEW OF THE LITERATURE

In recent years elementary school science has been involved in a profound change in philosophical approach. The basic emphasis in science teaching has been shifting from a content to a process orientation. Many educational organizations have tried to meet the challenge of introducing students to the processes of science by developing new curricula which stress the inquiry approach to learning science. Such new curricula include the AAAS Science--A Process Approach, the Elementary Science Study, the Science Curriculum Improvement Study, and the Elementary School Science Project. In view of the innumerable curricula available, decisions must be made as to which best meet their objectives and are worthwhile from the standpoint of implementation, in whole or in part. In order to make these decisions, some means of evaluation must be developed. "It is well to aim for new outcomes in science teaching, but we must also consider how we shall know whether we are achieving them (Munson, 1967, p. 127)."

THE EVALUATION OF PROCESS SKILLS

Welch and Walberg (1968) designed an evaluation guide for science curricula which indicates that a program should be evaluated on the basis of student gains in: (1) achievement

(content), (2) understanding methods of science and scientists, (3) interest in science, (4) attitudes towards subject area, and (5) knowledge of the processes of science. Many people have evaluated the new inquiry based programs with respect to content achievement by students, but the results of such evaluations have been inconclusive. For example, Carpenter (1963) found that students in an inquiry based program learned factual matter more readily than "traditional" students, whereas Bennett (1965) found that "inquiry" students did significantly poorer on a comprehension test and that no significant difference existed between the "inquiry" and "traditional" groups in factual knowledge.

The major problem which faces us, however, is evaluation of programs with respect to process skills, since no completely validated instrument for measuring process skills exists at this time. At present most of the support for the inquiry approach in science teaching is in the form of testimonial evidence. J. S. Bruner (1961) states:

> It is my hunch that it is only through the exercise of problem solving and the effort of discovery that one learns the working heuristic of discovery, and the more one has practice, the more likely is one to generalize what one has learned into a style of problem solving or inquiry that serves for any kind of task that one may encounter--or almost any kind of task. . . . Of only one thing I am convinced. I have never seen anybody improve in the art and technique of inquiry by any means other than engaging in inquiry (p. 31).

Nevertheless, there has been considerable effort expended to obtain experimental evidence concerning the acquisition of process skills. Peter Slinn (1969), using the STEP test, found significant differences in process skill acquisition in

two groups, both of which were using the inquiry approach to science. This serves as a reminder of the importance of the teacher variable in any program evaluation. Suchman (1960) and Butts and Jones (1966) found that scores on a test of problem solving ability were enhanced in students exposed to inquiry methods of teaching science. Scandura (1963) and de Tornyay (1967), on the other hand, found no significant differences in problem solving scores between "inquiry" and "traditional" students. Gagné (1962) found that his hypotheses concerning the hierarchy of process skills were supported by his experimental findings. He hypothesized that:

> (a) no individual could perform the final task without having these subordinate capabilities (i.e. without being able to perform these simpler and more general tasks); and (b) that any superordinate task in the hierarchy could be performed by an individual provided suitable instructions were given and provided the relevant subordinate knowledges could be recalled by him (p. 356).

The AAAS Commission on Science Education then quickly developed a curriculum called _Science--A Process Approach_ (see AAAS _Commentary for Teachers_, 1968) based on Gagné's process skill hierarchy. Their self-evaluation stated:

> . . . the behaviors designated by the authors of these exercises can be acquired by the intended audience of children (p. 73) . . . children exposed to the _Science--A Process Approach_ instructional materials acquire behaviors not already present in children who have not been exposed to these materials (AAAS _An Evaluation Model_, 1965, p. 57).

The commission's success in developing an instrument to measure the overall level of process skills in children has been somewhat slower. Correspondence with John R. Mayor (director of education for AAAS) revealed that the commission

is currently testing, but has not yet validated a science process instrument (SPI) designed to test process skills in grades four to six.

The question to be asked at this time is: How do we measure process skills? The analysis of standard science tests, such as STEP and TOUS, has shown a low weight on problem solving and reasoning and a high verbal factor (Hukins, 1963; Mokosch, 1969). Other more recently devised tests, such as the Test of Science Processes (Tannenbaum, 1969) have yet to be validated. It is no mean feat to validate a paper and pencil test designed to measure a behavioral skill. As Lisonbee (1966) warned, it is extremely difficult to make a test measure exactly what we want it to measure. We are just beginning to realize how little we know about testing.

Lacking a validated process instrument, the suggestion has been made that teachers should evaluate students in process achievement by keeping anecdotal records and checklists. Even this suggested method is not without its problems. As Reiner (1966) states:

> A challenge to devise such observational techniques that are simple, economical of time, practical and yield useful data, has not been satisfactorily answered up to now. The need is great (p. 336).

Thus, the inconclusiveness of the existing evidence on the acquisition of process skills and the measurement and evaluation of this acquisition points to the need for a great deal of further study in this area.

PIAGETIAN THEORY OF CLASSIFICATION

The importance of the process skill of classification has been well recognized by psychologists for over thirty years. This point was well emphasized by Bruner when he stated: "The learning and utilization of categories represents one of the most elementary and general forms of cognition by which man adjusts to his environment (Bruner, Goodnow, & Austin, 1956, p. 2)." Bruner further went on to break down this important skill into five distinct areas. He explained that the achievement of categorization would enable the organism to: (1) reduce the complexity of its environment, (2) identify the objects of its surrounding world, (3) reduce the necessity of constant learning, (4) provide direction for instrumental activity, and (5) order and relate classes of events. These achievements are only a start, however, as Bruner added,

> The moment an object is placed in a category, we have opened up a whole vista of possibilities for "going beyond" the category by virtue of the superordinate and causal relationships linking this category to others (Bruner, Goodnow, & Austin, 1956, p. 13).

Characteristic Criteria of Additive Classification

Piaget's definition of classification is as follows: "To find the common element in two given classes is to construct the relations between the given individuals and through this construction to reach a classification (Piaget, 1928, p. 196)." Piaget, in his detailed work on the development of classification in children (Inhelder & Piaget, 1964, p. 48), has outlined the criteria which are characteristic of what he con-

siders to be true additive classification.

1. The first requirement of classification is that there must be no isolated elements left over after a group of elements has been classified. That is to say, all the elements must be classified. Even if an element is the only one of its kind it must give rise to its own specific (but singular) class.

2. There must be no isolated classes. For every class A characterized by property "a" there must exist its complementary class A' characterized by the property "not a" (-a), such that A+A'=B.

3. The class A must include <u>all</u> those elements having the property "a". This is known as the extension of the class.

4. The class A must include <u>only</u> those elements having the property "a". This is known as the intension of the class.

5. All classes of the same rank must be disjoint. That is, no element may belong to two classes of the same rank simultaneously. AxA'=O.

6. A complementary class A' has its own characteristics which are not possessed by its complement A.

7. Any class A is included in every higher ranking class which contains all its elements plus other elements. For example, class B may contain all of A plus A'; thus yielding the following relations: B=A+A', A=B-A', and AxB=A, which means "all" the A are "some" of the B. This characteristic is the necessary requirement for class inclusion which indicates true classification.

8. Beyond the true class inclusion of paragraph

seven above, there is an effort toward extensional simplicity in which the inclusions are reduced to the minimim compatible with the intensional properties (i.e. try to make the fewest possible piles).

9. Intensional simplicity, or use of similar criteria (eg. colors) which distinguish classes of the same rank, also becomes a goal.

10. Finally one finds symmetrical subdivision. If a class B_1 is subdivided into A_1 and A_1', and the same criterion is applicable to B_2, then B_2 must likewise be subdivided into A_2 and A_2'.

On the basis of these ten criteria Piaget describes three stages in the development of classification ability in children.

Graphic Collections

Stage I in this development is known as the stage of graphic collections, and is usually found in children between the ages of two to five years. This stage of development does not universally involve any of the ten characteristics of true classification. The term "graphic collection" refers to a spatial arrangement of the elements classified, when it appears that the configuration of the elements played an essential part in the eyes of the subject. Stage I classification, then, is characterized by perceptual configurations based on the shape of the whole, with no understanding of any need for the operation of a characteristic similarity

within the class.

Piaget described two properties which are necessary and sufficient to make a class; the intensive properties:

> (1) (a) Properties which are common to the members of the given class and those of other classes to which it belongs.
> (b) Properties which are specific to the members of the given class and which differentiate them from members of other classes (Inhelder & Piaget, 1964, p. 17),

and the extensive properties:

> (2) Part-whole relations of class-membership and inclusion. These are conveyed by the quantifiers "all", "some" (including "one") and "none", when applied to the members of the given class and to those of the classes to which it belongs, insofar as they are qualified under (1a) and (1b) (Inhelder & Piaget, 1964, p. 17).

These two properties are always in correspondence in true classification so that whenever one is known, the other can be determined. However, in Stage I this correspondence does not exist. The children are perfectly able to discover relations of similarity and difference by successive assimilation but they frequently lapse into other forms of association.

> In particular they are constantly misled by consideration of pattern or by the situational and descriptive properties of the material. Even more significant is the fact that because these assimilations are only successive they cannot quantify their result (Inhelder & Piaget, 1964, p. 45).

Thus, successive assimilations impose a limitation on the relationships of similarity and difference at the level of graphic collections. These relationships are only applied to objects in successive pairs and remain unconnected with part-whole relationships. "The graphic collection [then] stands midway between a composite spatial object and a class

(Inhelder & Piaget, 1964, p. 19)."

The problem that the Stage I child has with the meanings of "some" and "all" is related to these difficulties. Although the concepts related with these words exist in embryo they are not fully differentiated nor completely co-ordinated with one another. The word "all" to the Stage I child often admits exceptions while the word "some" occasionally includes "all". Class-inclusion (item seven on the criterion scale, page 14) is far beyond the abilities of the Stage I child, and even class-membership cannot be achieved until there is a simultaneous awareness of the similarities and differences of the objects involved.

There does not appear to be any definite sequence of sub-stages within the stage of graphic collections. However, certain recognizable, overlapping reactions do occur. Piaget describes three main types of graphic collections. The first involves "laying the objects in line; the line may be continuous or discontinuous (Inhelder & Piaget, 1964, p. 20)." For example, the child may start by putting two objects together which have a similar shape. The third object in the line, however, will not necessarily have the same shape as the others. Often it will be similar to the second of the two previous objects in some other property, say color. This activity continues, with each object being chosen on the basis of similarity with any one of a number of the properties of the preceding object, until all the objects belong to a continuous or discontinuous line.

A second type of graphic collection is that of collective objects: ". . . graphic collections in two or three dimensions made up of similar elements, so that together they constitute an unbroken whole, this being a geometrical figure or pattern (Inhelder & Piaget, 1964, p. 20)." This collection may easily be mistaken for a more sophisticated categorization unless one is sure to ask about the reasons for the grouping. The reason given may make it clear that the child has not isolated the intensive and extensive properties of the class, but has merely used similar elements to facilitate making the "whole".

Finally, one finds the occurrence of complex objects: ". . . unlike collective objects, these are made up out of heterogeneous elements. The end-product may be a geometrical figure, but it can also be descriptive or pictorial (Inhelder & Piaget, 1964, p. 20)." Rawson (1965, p. 45) describes this type of pre-classification behavior as extending well into elementary school. It may consist of notions of "belonging together" such as putting a "mother" with a "baby" and a "bath" because together they tell a story. The child visualizes a construction not a class, and usually defines words by their use rather than by their meaning.

In terms of flexibility in classifying, the Stage I child also exhibits limitations. Once one criterion has been established it is adhered to very rigidly. Piaget calls this phenomenon "perseveration". After a time, however, the child often forgets what his criterion was and is thus enabled to

introduce another criterion. This process usually ends with the child remembering and reverting back to his original criterion. Occasionally several criteria are thought of simultaneously and mixed together, thus producing a very arbitrary grouping such as: "They're better over there".

In summing up Piaget considers ". . . one key to the problems of stage I is failure to abstract (Inhelder & Piaget, 1964, p. 79)." The results of his experiments on classification bring him to the conclusion that:

> There is a stage of the graphic collection, which may be long or short depending on the materials used and the instructions given. This stage invariably precedes the stage of the non-graphic collection (this being a collection based on similarity and difference alone, so that class-membership is apparent while class-inclusion is not). True classification is later than either of these (Inhelder & Piaget, 1964, p. 20).

Non-Graphic Collections

Stage II is described as the stage of non-graphic collections and is usually found in children between the ages of five to seven or eight years. It is largely a transition stage in which a steady growth in the number of criteria for classification takes place. Criteria one, two, three, and four (p. 14) are almost always achieved and five and six are often considered. Even criteria eight, nine, and ten occasionally occur. This growth, however, always falls short of achieving class-inclusion, the number seven criterion, and consequently does not reach the point of true classification. "However, these collections are no longer graphic, and objects are assigned to one collection or another on the basis of

similarity alone (Inhelder & Piaget, 1964, p. 47)." The growth of the non-graphic collection stage may be attributed largely to the increasing differentiation of intension and extension, and hence, a greater co-ordination of the two. The development of the meanings of "all" and "some" show considerable improvement in this stage although "all" still may be confused with "many" and "some" with "few". In the event of a small collection "some" frequently merges into "all". Piaget cites the following example of a Stage II child:

> Thi (6;7) correctly reproduces the row of blue circles and blue and red squares from memory. "Are all the red ones squares?--<u>No, because there are blue ones too</u> (wrong).--Are all the blue ones circles?--<u>Yes</u> (wrong).--Are all the circles blue?--<u>Yes</u>-- Are all the squares red?--<u>Yes, with two blue squares</u> (!)" (Inhelder & Piaget, 1964, p. 62).

The Stage II child often wrongly takes the statement "All the A's are B's" to mean "All the A's are all the B's". This equates A with B (A=B) which is an <u>identity</u> relationship rather than the correct relationship of <u>inclusion</u>. The inclusion relationship must be "All the A's are some of the B's" which implies that B is greater than A (B>A). The Stage II child also equates the question "Are all the blue ones heavy?" with the question "Are all the heavy ones blue?". He will answer the first form of the question no matter which form is asked because he finds it easier to think about a collection defined by color rather than by weight. In general, at Stage II, children have developed some differentiation between the meaning of the words "all" and "some" but they

still fail to grasp the essentially relative meaning of "some". ". . . even in the second half of stage II, 'some' has an absolute meaning which is bound up with the number of elements instead of pointing to a relation between part and whole (Inhelder & Piaget, 1964, p. 93)."

Like the graphic collection, the non-graphic collection is bound by a condition of spatial proximity. To reach Stage III, class-inclusion must become the alternative for spatial proximity. Piaget draws the distinction between the subdivision of collections and class-inclusion in the strict sense, even if the distinction is not always easy. The difference is that the general group B continues to exist even when it is subdivided into A and A', in true inclusion. The child is then able to reason in the form A=B-A'. In the non-graphic stage the child looses sight of B when A and A' are dissociated. For example, if you were to present a Stage II child with twelve wooden beads, two of which were red and ten yellow, and asked him to state whether there were more wooden beads or more yellow beads, he is likely to answer "yellow". The predominance of the yellow beads (A) reduces the wooden beads (B) to the same level of importance as the red beads (A'). In other words the child is said to have inclusion only if he truly understands that "all" the A are "some" of the B.

Piaget outlines four main types of response to a classification task that occur in Stage II, the first of which is characterized by unstable criteria and by the assembly of

collections that do not exhaust the objects. Each component of the total response is a reaction to an immediate and partial problem. This type of response involves no overall plan. As the child progresses to a "type two" response he revises his criteria in order to use up all the objects. At first this process involves trial and error but it is soon aided by hindsight (retroaction). The "type three" response sees this retroaction turn to partial anticipations which bring about the emergence of a principal criterion. After a time, the collections so arrived at may be further subdivided, thus characterizing the fourth type of response possible in Stage II.

The key factors in the growth of these stages seem to be the gradual adjustment to trial and error grouping, with an accompanying increase in retroaction and anticipation. However, these children are not beyond making mistakes and showing considerable confusion. There is relatively less forgetting and perseveration than in Stage I, but these reactions become more common when trying a third or fourth criterion. In such instances, mixed and contradictory collections are still occasionally found. Nevertheless, Piaget reports:

> . . . what these results do show clearly is that there is a retroactive process whose influence increases. We see it in the various rearrangements which follow on the introduction of new elements (Inhelder & Piaget, 1964, p. 205).

We must remember, however, that hindsight does not insure flexibility. Stage II children use an ascending method of classifying. They begin by studying each of the shapes one

at a time and make a number of small collections before they discover any true dichotomies. They usually stick rigidly to the first dichotomy they find. No systematic survey of the material is made nor are any plans formulated. There is still little evidence of anticipation.

The main problem of the Stage II child is that of ego-centrism. He experiences difficulty in becoming aware of his own thought and consequently reasons only about isolated or particular cases. Generalization does not come naturally to him. For example, he has difficulty dissociating the notions of "going together" and being "well arranged". A child at this stage also has to cope with two planes of reality--that of the play world and the reality of ordinary life. These planes, for the Stage II child, are in a juxtaposed position instead of in an ordered hierarchy as in Stage III. As Piaget (1928) explained:

> But as regards the verbal faculty, these different planes have not yet come to be distinguished, so that the child is incapable of conceiving on the one hand of a purely logical necessity (If we say that . . . then we must say that . . .), on the other hand of a plane inhabited by pure hypothesis or by logical assumption (Let us say that . . .) (p. 250).

The problem of ego-centrism further influences the child's actual perceptions of his environment.

> He does not analyse the contents of his perceptions, but weighs it down with a load of previously acquired and ill-digested material. . . . The child often sees only what he already knows (Piaget, 1928, pp. 248-249).

The non-graphic collection, then, can be described as a mental "pile" or heap of things, the shape of which is

irrelevant. However, it is still just a "collection" and not a "class" because the constituent elements must be perceptible and fairly close together. They are held together only by a qualitative criterion and they lack reversible mobility.

True Classification

The final stage that Piaget talks about might be described as the stage of true classification. This stage usually occurs sometime after the child's eighth year. It is characterized by the ability to consider all ten of the criteria of classification, even the seventh criterion class-inclusion. The first indication of the onset of this stage is a change in the logical judgment of the child. As Piaget (1928) states:

> . . . about the age of 7-8, certain changes take place in the modality of childish judgment which are in close relation to the appearance of the desire for a system and non-contradiction (p. 249).

This desire is facilitated by the transformation of the planes of reality into a hierarchical order. Observations of phenomena become bound by physical and not moral necessity. When children in Stage III are presented with new elements for an existing classification, they are capable of fitting them into the original framework of criteria correctly, or they restructure their criteria. They have overcome the problems of intension and extension. As McCormack (1969) illustrated these problems in Stage II:

> These difficulties with extension and intension come from irreversibility of thought. Once a child has conceived of a set of objects as being in one class (or

> possessing one property) he cannot return to the starting point of his mental construction and reconceive the same set of objects as being in a second class (or possessing a second property) (p. 24).

These Stage III children have also achieved a clear differentiation between "some" and "all", and have a thorough understanding of class-inclusion. Even a student who forms three or even two large collections and then divides each of them into sub-collections has not yet demonstrated the attainment of Stage III. He has shown the relations ($B_1=A_1+A_1'$) and ($B_2=A_2+A_2'$), but he must still demonstrate the relation of true inclusion ($A=B-A'$). This proves his understanding of "all" and "some" and the relation ($B>A$).

Piaget gives, as an example of Stage III, a conversation with Cor (age six years, eight months), who had just correctly reproduced the row of blue circles and blue and red squares from memory:

> Are all the red ones squares?--Yes.--You are sure?--Yes.--Are all the blue ones circles?--No, not all. There are also [blue] squares.--Are all the squares blue?--No, there are also red ones.--Are all the circles blue?--Yes. (Inhelder & Piaget, 1964, p. 63)

Cor is typical of Stage III and answers the questions with no mistakes and considerable confidence.

Flexibility, now, becomes a more important part of the child's classifying ability. No perseveration is observed at this stage and little or no forgetting of previous criteria occurs. The child is capable of directly picking out the most general characters and then subdividing them into the major classes. This illustrates the skill of anticipating

several dichotomies. As Piaget explains it, "All this suggests that the anticipatory flexibility is what accounts for the flexibility in hindsight when a change of intension is called for (Inhelder & Piaget, 1964, p. 215)." It must be remembered that the transition from the assembling of graphic collections to that of true classification is a very gradual one. The division of this process into three stages is by no means an attempt to show otherwise. Stage III behavior is simply the final phase in a process of continuous development. Piaget emphasizes that his studies on both the logic of relations and the logic of classes confirm this gradual development. As he puts it:

> Both of them show that the thought of the child passes from a state of ego-centric immediacy, in which single objects only are known and thought of absolutely, and made to bear no relation to one another, to a state of objective relativism in which the mind extracts from these objects innumerable relations capable of bringing about the generalization of propositions and reciprocity of different points of view (Piaget, 1928, pp. 133-134).

RESEARCH RELATED TO CLASSIFICATION

Many studies have been made of the number and kinds of classifications that children make, not the least of which are the experiments that Piaget himself has carried out. In his study of flexibility, Piaget attempted to determine how well children could change their criteria of classification. He presented children, aged five to nine, with a group of cardboard cut-outs differing in color, shape, and size. They were asked to classify the cut-outs into two groups. Those

who completed one classification were asked to classify the cut-outs a different way, and then to attempt a third way. The following table is a presentation of the results of this experiment:

Number of criteria at ages 5-9

Age (no. of subjects)		5 (12) %	6 (17) %	7 (18) %	8-9 (13) %
Criteria:	0	27	12	5	0
	1	46	12	11	0
	2	27	47	56	31
	3	0	29	28	69
	2 or 3	27	76	84	100

(Inhelder & Piaget, 1964, p. 209)

Piaget indicated that it was significant that over 75 percent of the six year olds classified two or three times. He went on to say that;

> . . . once a child can divide the same objects according to two or three complete dichotomies, it is but a short step to being able to cross-classify them in accordance with a multiplicative schema (Inhelder & Piaget, 1964, p. 209).

For the purposes of this study, however, it seems significant that 100 percent of the eight to nine year olds completed two or three classifications. Does this result mean that these children are at their peak in classification performance and no further improvement can be expected of them, or does it mean that a test which allows only three dichotomous classifications does not have enough scope to discriminate between the differences of classification ability in these children?

Piaget went on to observe more closely only those

children who were able to make at least one classification. He found that 88 percent of these seven year olds were able to make one or two changes in their classifications and that no significant improvement in this percentage took place up to and including the age of nine years (which was the age limit for his sample). This evidence would also seem to indicate that the peak in classification ability occurs at about the age of seven or eight.

Reichard's evidence also confirmed this conclusion. He tested 234 normal white children aged four to fourteen, using the Weigl, Goldstein, and Scheerer color form test (see Goldstein & Scheerer, 1941) and a sorting test composed of thirty-three items. Reichard found that below five years of age most children grouped once and then were unable to shift from their original criterion. At the age of seven to eight two groupings became most common. Over eight years of age 75 percent of the children could shift from one criterion to another. On the sorting test Reichard found a steady increase in the number of correct responses up to the age of nine. As he explained:

> The rate of this increase appears to approximate the normal growth curve, and shows decreasing acceleration which reaches a plateau at about 9 years, after which there is almost no further increase (Reichard, Schneider, & Rapaport, 1944, p. 158).

Reichard also noted from the sorting that the most common criteria for sorting were those of use, color, form, material, and matching pairs.

Kofsky (1966) carried out a similar grouping task with

children aged four to nine and found a significant correlation between age and the number of tasks mastered. However, contrary to Piaget and Reichard, he found no indication that this trend was leveling off at the age of nine. For more detailed information regarding this discrepancy, the results of tests on children above nine years of age must be examined. Croxton (1963) presented children from kindergarten to grade eight with a table on which was placed a large group of plants on one side and a variety of animals on the other side. He then asked each child to state as many ways as possible in which all of the animals on the table were different from all of the plants. He found that the range and the number of generalizations gradually increased throughout the grades. Croxton concluded that:

1. Children in the higher primary grades and beyond are capable of generalizing.

2. Early childhood is pre-eminently a period for satisfying interactions (i.e. gaining experience with objects in the environment without thought of classifying them).

3. Junior high school students (grades seven and eight) do little better than upper elementary students.

Additional information has been provided by Thompson (1941), who tested the classification ability of children from grades one to six. She used three different tests to obtain her data: the Weigl color form test (see Goldstein & Scheerer, 1941); the BRL test (see Bolles, 1937); and the Vigostsky test of concept formation (see Hanfmann & Kasanin, 1937). Thompson

divided the students into two groups, the grade four, five, and six group and grade one, two, and three group. On the Weigl color form test, which allowed for the possibility of making zero, one, or two classifications, the older group completed two classifications significantly more times than the younger group. They also demonstrated a greater ability to generalize. The results of the BRL test, which consists of a collection of thirty-four objects to be classified, yielded 6.3 average classifications for the older group and 4.1 average classifications for the younger group, although a wide range of variability was found in both groups. Thompson concluded that:

> . . . there is a definite qualitative and quantitative change in the ability to generalize or to handle abstractions between the age range 6 to 8 years, and 9 to 11 years (Thompson, 1941, p. 124).

With regard to the kinds of classifications children make, most studies have limited themselves to testing the child's preference for form or color. Thompson (1941) found that both of her groups of children gave significantly more color generalizations than any other kind. The younger group found color by far the easiest criterion to use, whereas the older group found some of the other physical properties to be just as easy to identify and use as a criterion. Corah (1964), on the other hand, found that the children she tested shifted from color to form preference at about the age of six. Suchman and Trabasso (1966) also differed in their conclusions from Thompson in that their results indicated that children

below the age of 4.2 years usually preferred color, while children from age 4.2 to 6 years usually preferred form. Thompson's findings also contradict the long held notion that the transition from color to form preference is a correlate of cognitive growth (Suchman and Trabasso, 1966). This latter notion is supported by the fact that color responses to the Rorschach test are termed "primitive" or "immature" (Schactel, 1959). Also, Brian and Goodenough (1929) found that "form preferring" children obtain higher mental test scores than "color preferring" children, and "form" children are more accurate in classifying stimuli along other dimensions (Suchman, 1966).

Kagan and Lemkin (1961) also studied form and color preference in children and found what seemed to be a sex difference as well as an age difference in this preference. They reported that although both boys and girls preferred form to color or size, older boys (aged eight) used color more often than similarly aged girls, and that the eight year old girls used form more than the younger girls.

A number of researchers have categorized the criteria of classifications into several general types. Bruner noted that;

> One may distinguish three broad classes of equivalence categories, each one distinguished by the kind of defining response involved. They may be called affective, functional, and formal categories (Bruner, Goodnow, & Austin, 1956, p. 4).

Honkavaara (1958) worked with similar categories which he called dynamic affective, concrete, and abstract respectively.

He found that bright children in the eight to nine year old range categorized on an abstract level whereas backward children of the same age were much more concrete. However, when younger children were tested, backward children increased in affective groupings while bright children became more concrete, but not affective.

Thompson (1941) divided the classifications her subjects made into five categories: (1) general name, (2) common use, (3) concrete situation, (4) common attributes, and (5) no explanation. She found that young children, four to six years old, were very concrete in their classifications while the older children relied on usage and the general name. Annett (1959) tested the classification abilities of children aged five to eleven (and some adults) with a card sort consisting of sixteen cards, divided equally into the categories of vehicles, animals, plants, and furniture. Like Thompson, Annett found that the extremes of her five categories were "no explanation" and "class name". However, Annett described the middle three categories as: enumeration, contiguity, and similarity.

Attacking the problem from a different angle, Sigel (1954) wondered if the level of symbolism or the degree of manipulation affected the ability to classify. He set up a battery of five tests designed to answer this question. Test one allowed children to manipulate twenty-four toys representative of common objects such as tractors and horses. Test two presented children with the same set of toys to classify

but did not permit any manipulation. The third test replaced the toys with black and white photographs of the same objects and allowed these pictures to be manipulated. Test four replaced these pictures with the names of the objects, each name being printed in big block letters on separate pieces of paper. The fifth test presented the child with the names of the objects listed on one sheet of paper in regular type. The child was required to write on another page the names of the objects he wanted to put together. The directions given for these tests (except number five) were:

> Put those things together in one pile that belong together or go together or are alike in any way, and those other things that go together or belong together or are alike in another way in another pile. You may have as many piles or as few piles as you wish. Do you understand (Sigel, 1954, p. 202)?

Sigel then proceeded to test students seven, nine, and eleven years of age. He designated the classifications obtained as either perceptual or conceptual. The number of classifications in each category provided the basis of a quantitative score. Sigel found no significant differences in either the perceptual or the conceptual test scores among any of the five presentations at each of the three age levels. He did note, however, that most of the children made their groupings primarily on use. They seemed to ignore stimulus characteristics like color, texture, size, or material. Sigel called this phenomenon "meaning dominance in classification".

To sum up, the results of the studies reported here reveal some disagreement as to the growth of classification

ability in students beyond nine years of age. There is no consensus of opinion, either, on the significance of preference for color, form, shape, or use as criteria for classifying. Even the general types of criteria used differ from one report to another. Thus it is indicated that further research on the classification abilities of elementary school children may prove fruitful.

Chapter 3

DESIGN OF THE STUDY

THE SAMPLE

The population consisted of all of the elementary students from twelve different schools, six of which were in the Edmonton Public School System and the other six within the County of Strathcona #20. These schools were selected by representatives of the two systems involved to give a sample of students with a broad range of socio-economic status and cultural background. A sample of one hundred and eighty children was selected from these schools in such a way that only fifteen students were used from each school. The fifteen students from each school included two or three children from each of the grades one to six. The number of children to be taken from each grade in each of the schools was determined beforehand so that there would be a total from all schools of thirty students for each grade when the testing was finished. All the students were randomly selected from class lists of the proper grades. This randomizing procedure was carried out to minimize the effect of any particular teacher or school on the results of the study. Table I summarizes the sample with respect to boys and girls in each grade.

TABLE I

NUMBER OF BOYS AND GIRLS PER GRADE

Grade	1	2	3	4	5	6	Total
Boys	17	14	13	13	16	16	89
Girls	13	16	17	17	14	14	91
Total	30	30	30	30	30	30	180

INSTRUMENTATION

Classification Task

Theoretical background. In order to achieve the purposes of this study, an individualized classification activity was designed. This activity, while meeting the definition of classification described in Chapter 1, had to meet other more common conceptions of classification. A typical Piagetian definition of classification is described by Rawson (1965):

> A class is a conceptual judgment arrived at by listing all the members of the class (the extension of the class) and by naming the property shared by every member of the class and no other (the intensional property of the class) (p. 51).

From a natural scientist's point of view there are other important understandings of classification: (a) criteria for classifying are arbitrarily chosen, (b) the determination of the class to which an object is most appropriately assigned involves a matter of judgment, and (c) the number of criteria

found useful in grouping is a measure of classification ability (Unusual Objects, 1967, pp. 86-87).

One must also consider Piaget's ten criteria for classification as outlined in Chapter 2 (pp. 14-15). Testing separately for the presence of each criterion would prove very time consuming. What was needed was a classification activity that would be representative of these ten criteria and would meet the requirements of the other definitions and understandings of classification as well. In addition, this classification activity had to have more discriminative power than is afforded by only three possible criteria, and should also encompass the scope necessary for the testing of children from grades one to six.

McCormack (1969) hinted that flexibility in handling elements when classifying was a good overall measure of classification ability. Confirmation of this was found in Inhelder and Piaget (1964):

> The main difference between the operational classifications of stage III and the imaginal or graphic classifications found at stage I is that the child who is more mature is very much more flexible in the way he handles the elements. This applies both to his mental perspective and to the actual shuffling and reshuffling to which he is capable. He shows flexibility in hindsight when he can change the criterion, either because he notices some property which he failed to take into account when he started, or because the experimenter brings in additional elements to add to an existing classification. He shows flexibility in foresight when he can mentally anticipate a classification before putting it into practice, and particularly when he is able to choose the best classification out of a number of possible alternatives without overt trial-and-error (p. 196).

It was decided, on the basis of this statement, to use an

activity which tested flexibility, or the ability to change criteria when a significant number of alternative classification criteria are possible.

Materials. The materials used in the experimental classification task consisted of twelve different substances. Each of which occupied up to one and one-quarter to two inches of space in numbered, capped, square bottles. The substances consisted of four liquids (orange colored water, corn syrup, mercury, and glycol) and eight solids. The solids fell into three different categories: chips, powders and crystals. The chips were aluminum and copper; the powders were flour, powdered aluminum, sulphur, and instant coffee; and crystals were potassium dichromate (orange) and white sugar.

To facilitate the preparation of students for their task, a set of eleven white cardboard shapes was used for preliminary instruction. This set included three large (three inch) squares, three small (one and one-half inch) squares, two large circles (three inch diameter), two small circles (one and one-half inch diameter), and an equilateral triangle (three inch base).

A data sheet was used to record the objects the child grouped together in each classification, as well as other information about the child (see Appendix B, page 89).

Procedure. In order to minimize the verbal aspect of the presentation of procedural directions and to ensure an adequate understanding of the task to be performed (these two factors being especially important in the lower grades), each

child was shown the set of cardboard cut-outs. They were grouped (by the investigator) first according to shape and then according to size. The child was asked to guess the criterion for grouping in each case. After this procedure was completed, attention was focussed on the bottles the child was then to classify. The directions given to the child were similar to those used by Inhelder and Piaget (1964) when they studied classification and the relative size of classes. They explain:

> We then proceed through the following questions: (1) the subject is told to classify these objects as he likes; (1b) if he has not done so already, he is told to divide the objects into two classes only; (2) he is asked to redivide them into two classes using a different criterion. (3); he is urged to do the same again, using a third criterion; (p. 125) . . . The most usual instruction was, "Put together things that are alike." But not infrequently this was further amplified; e.g. ". . . Put them here if they're the same, and then over there if they're another lot different from this one but the same as each other (p. 21)."

However, some differences from these instructions did exist in the present investigation. Step (1b) of Inhelder and Piaget's test was omitted because many of the possible criteria for classification did not naturally present a dichotomous situation. The child, in addition, was urged to use as many different criteria he could possibly think of. If the child did not use any other sense than sight for his classifications, he was asked: "If you could take the lids off the bottles and do anything you wanted to do with the materials inside, how else might you be able to group them?"

At the completion of each classification, the child

was asked to explain why the things he put in each group belonged together. In order to determine the time element involved in the final score, a stop watch was started the first five times the child began to think of a way to group the bottles, and was stopped immediately upon the completion of each grouping. Although the time was recorded, no time limit was set for the completion of the total activity. However, the test was concluded when a child failed to attempt a new classification within a period of five minutes from the conclusion of the completion and explanation of his previous classification.

Scoring. As the child classified the material in the bottles, a record, by bottle number, of the bottles placed into each group was kept. At the end of every classification a judgment was made, on the basis of the groups formed and the verbal explanation of the child, as to what the criterion of the classification was. This criterion was indicated alongside the number groups recorded on the data sheet. The main task essentially required only successive linear classifications. If a simultaneous or multiplicative classification was made, each property isolated was credited as a single criterion and the child received the same number of points as he would have if he had made each of the classifications separately. Each different criterion of classification was then given a base value of ten points (classifications based on external marks or deformities on the bottles themselves were not given any value, only those criteria referring

directly to the substances in the bottles were evaluated). However, if the child made any obvious errors in grouping according to a criterion, or was unable to exhaust the elements (bottled samples) in making the classification, one point was subtracted from the base value for each element misplaced or not used. The sum of these criterion values thus obtained is the power score as defined for this study. This score reflects the number of classifications performed minus the errors or ommissions made in grouping.

In an effort to add more discriminative power to the test, a time factor was introduced. The time, measured from the moment the child started to think about a way to group the bottles until he had completed putting the bottles into their groups, was recorded for the first five classifications made. Only the first five classifications were timed in order to prevent unfair discrimination against children who were overly persistent in searching for more criteria after they had used all those that were more immediately apparent to them. The speed score was determined by computing the arithmetic mean of the five recorded time intervals. A composite score was needed in order to relate the time score to the power score. This composite score was obtained by dividing the power score by the speed score. Thus a child who received the same power score as another child, but took half the average time to complete the classifications, would receive a composite score twice that of the other child.

I.Q. Scores

No I.Q. tests were administered by the investigator during this study. I.Q. scores for the members of the sample were obtained from school cumulative records. Unfortunately, the I.Q. scores for students in any particular grade could have been obtained from one of three different I.Q. tests: the Alpha Short Form of the Otis Quick-Scoring Mental Ability Tests, the Detroit Beginning First-Grade Intelligence Test, or the Lorge-Thorndike Intelligence Test, Level 3, Verbal Battery. The reliability of the intelligence test scores assigned was further reduced because of the fact that 18 percent of the students in the sample had no record of an I.Q. score throughout their school years. These students, for the purposes of this study, were placed in high, average, or low intelligence categories on the basis of recommendations made by their classroom teachers. The circumstances described made it impossible to deal with I.Q. on any basis other than three broad categories. On the basis of a suggestion made in the Detroit Beginning First-Grade Intelligence Test Manual (Engel & Baker, 1937, p. 7), the top and bottom 20 percent of the students in each grade were assigned to the high and low intelligence categories respectively. The middle 60 percent of the students were categorized as the average group.

THE TESTING PROGRAM

The testing program was carried out during the last week of April and the first three weeks of May, 1970. The

investigator travelled to each of the twelve schools and administered the classification task individually to every member of the sample. The testing was done in a private room in each school (usually the medical room or the guidance counsellor's office) which was as free from distractions as possible. Appointments were made with each school to carry out testing on days that did not coincide with field trips, festivals, or other special events that may have eliminated a particular group of children from the sample; a situation which may have led to some bias in the final results.

THE PILOT STUDY

A pilot study was carried out in March, 1970, with twelve children who were not involved in the main study. This group of twelve children consisted of four children from each of the grades two, four, and six. The pilot study was undertaken to determine:

1. The approximate time needed for children to complete the classification activity.

2. The type and amount of instruction needed in order to have the task understood by children in each grade.

3. The feasibility of presenting the designed classification task to a broad range of grades.

4. The likelihood of finding significant differences in performance of this task throughout the grades tested.

5. Administrative problems that may have manifested themselves.

On the basis of the pilot study it was decided that:

1. The approximate time needed for completion of the classification task was fifteen minutes for the lower grades, extending to thirty minutes for many of the children in grade six. However, in most cases the motivation for continuing the task was high and no children gave the impression that they wanted to stop classifying because of boredom rather than because they could think of no more criteria. The persistence that many children showed in trying to think of more criteria made it clear that some time limit had to be established in order to prevent children from sitting for long periods of time without arriving at a new criterion and yet refusing to give up. The time limit of five minutes was judged as ample for most children to think of another criterion.

2. Completely verbal directions were not adequate nor suitable for all children in the pilot study. Verbally backward children were severely handicapped unless given extremely detailed instructions which tended to bore the more verbally adept students. It was found that classifying the contents of the bottles in one way as an example for the child saved much time and effort that otherwise would have been expended in verbal explanation and was more universally understood by the children. This method also made it much easier to be consistent in giving the instructions for the task. However, this method used up one criterion that some of the children may have discovered for themselves. To avoid this problem,

the classification example was given instead by means of a group of white cardboard cut-outs that were thought to have little transfer effect (with regard to classification criteria) with the bottle contents.

3. With the above refinements, children from the three grades tested were able to attempt the classification task.

4. A steady increase in the number of classifications occurred as the grade level rose. Girls made slightly more classifications than did boys. These results are summarized in Table II.

TABLE II

NUMBER OF CLASSIFICATIONS
PER GRADE AND SEX
(PILOT STUDY)

	Grade II	Grade IV	Grade VI	Boys	Girls
Number of Classifications	13	20	37	32	38

At least some members of each of the grades were able to classify by the criteria of: color, liquid vs. solid state, grain size, odour, touch, and lustre. However, only some of the grade sixes were able to make use of: cohesiveness, grain shape, transparency, viscosity, taste, sound, and use.

5. A private room furnished with a fairly large flat-topped table and two chairs facilitated the administration of the task most adequately. Care had to be taken in those rooms

containing windows to prevent direct sunlight from shining on the bottles as such a situation had the effect of making the property of transparency more obvious to the children.

THE TYPE OF ANALYSIS USED

Hypotheses one and two were subjected to a two-way analysis of variance to determine the significance of interaction (if any) between grade and I.Q. and grade and sex, respectively. This interaction was measured by a test for additivity carried out by means of the ANOV25 IBM 360/67 computer program. The 0.05 level of significance was used as a basis for the decision to accept or reject these two hypotheses. The remaining three hypotheses were subjected to a one-way analysis of variance. Hypothesis number three, comparing boys scores with girls scores, was analysed with a "t" test computed by use of the ANOV10 IBM 360/67 computer program. The decision regarding acceptance or rejection of this hypothesis was also made at the 0.05 level of significance. Hypotheses four and five were analysed by means of the ANOV15 IBM 360/67 computer program. This program carries out a Scheffé multiple comparison of means. The 0.10 level of significance was used as the criterion for accepting or rejecting these hypotheses. This rather low level of significance was chosen because of the extremely conservative nature of the Scheffé procedure. As Ferguson (1966) points out:

> Concern may attach to the fact that the Scheffé procedure

is more rigorous than other procedures, and will lead to fewer significant results. Because this is so, the investigator may choose to employ a less rigorous significance level in using the Scheffé procedure; that is, the .10 level may be used instead of the .05 level. This is Scheffé's recommendation (1959) (p. 297).

Chapter 4

RESULTS OF THE INVESTIGATION

The results of the statistical analysis for each of the hypotheses as well as a description of the number and kinds of criteria used in the various grades are given in this chapter. The report on the statistical analysis of hypotheses includes analysis of variance tables, cell means matrices, and probability matrices. These were calculated by means of the IBM 360/67 analysis of variance computer programs which were documented and tested by members of the Division of Educational Research Services of the University of Alberta.

STATISTICAL ANALYSIS OF THE HYPOTHESES

Hypothesis #1

There is no significant interaction between grade level and I.Q. with respect to:

(a) Classification power score,

(b) Classification speed score,

(c) Composite classification score.

The purpose of this hypothesis was to determine if I.Q. has any differential effect on each of the criterion scores in different grades. This hypothesis was analysed by means of a test for additivity calculated by utilization of

the ANOV25 IBM 360/67 computer program.

Classification power score. The mean power scores for each I.Q. group in each grade are presented in Table III, page 50. The test for additivity on this interaction yielded an F value of 1.653 which corresponds, in this case, to a level of significance of 0.096. The criterion level of significance was taken to be 0.05. Therefore, the interaction is not significant and hypothesis one, with respect to this variable, could not be rejected. Table IV, page 50, provides a summary of this information.

Classification speed score. The mean speed scores for each I.Q. group in each grade are presented in Table III, page 50. The test for additivity on the speed score variable resulted in an F value of 1.175. The level of significance resulting from this value is 0.312. Therefore, the interaction is not significant at the 0.05 level and hypothesis one, with respect to the speed variable, was not rejected. Table IV, page 50, presents this information.

Composite classification score. The mean composite scores for each I.Q. group in each grade are also presented in Table III, page 50. The test for additivity on this interaction yielded an F value of 2.759 which indicates a highly significant interaction ($p<0.01$). Therefore, hypothesis one, with respect to the composite score variable was rejected. Table IV, page 50, provides a summary of this information.

TABLE III

MEAN CLASSIFICATION SCORES FOR EACH I.Q. GROUP PER GRADE

Variable	I.Q. Group	N	Score in grade 1	2	3	4	5	6
Power scores	High	6	42.67	74.83	85.17	71.33	109.67	113.67
	Average	18	29.28	38.61	64.78	70.78	82.89	79.56
	Low	6	19.00	31.17	33.67	48.83	64.17	111.00
Speed scores	High	6	1.78	1.42	1.67	1.25	0.80	0.90
	Average	18	1.71	2.09	1.16	1.23	1.02	1.34
	Low	6	1.47	2.45	1.52	1.55	1.78	0.98
Composite scores	High	6	28.50	51.17	73.00	78.83	173.17	139.00
	Average	18	22.11	29.61	73.06	67.56	93.17	74.83
	Low	6	14.50	17.50	28.33	32.67	45.17	130.67

TABLE IV

INTERACTION BETWEEN GRADE AND I.Q. ON CLASSIFICATION SCORES

Variable	Source of Variance	S.S.	D.F.	M.S.	F	Prob.
Power scores	Interaction	13927.6	10	1392.76	1.653	0.096
	Error	136530	162	842.776		
Speed scores	Interaction	8.320	10	0.832	1.175	0.312
	Error	114.737	162	0.708		
Composite scores	Interaction	57969.8	10	5796.98	2.759	<u>0.004</u>
	Error	340331	162	2100.81		

Underlining indicates significance at the 0.01 level

Conclusion. Hypothesis number one, dealing with the interaction between grade and I.Q., could not be rejected with respect to the variables of power and speed but was rejected with respect to the composite score variable. However, the composite score is not independent of the power and speed scores but is simply a ratio of the two. The significance of the interaction resulting from the composite score, then, is simply the product of the two insignificant interactions related to the power and speed variables. Further investigation of this interaction showed that it was caused mainly by the extremely high score obtained by the grade six, low I.Q. group. This interaction is graphically illustrated in Figure 1, page 52.

Hypothesis #2

There is no significant interaction between grade level and sex with respect to:

(a) Classification power score,

(b) Classification speed score,

(c) Composite classification score.

The purpose of hypothesis number two was to determine if sex has any differential effect on each of the criterion scores in different grades. This hypothesis was also analysed by means of a test for additivity calculated by the ANOV25 IBM 360/67 computer program.

Classification power score. The mean power scores for boys and girls in each grade are presented in Table V,

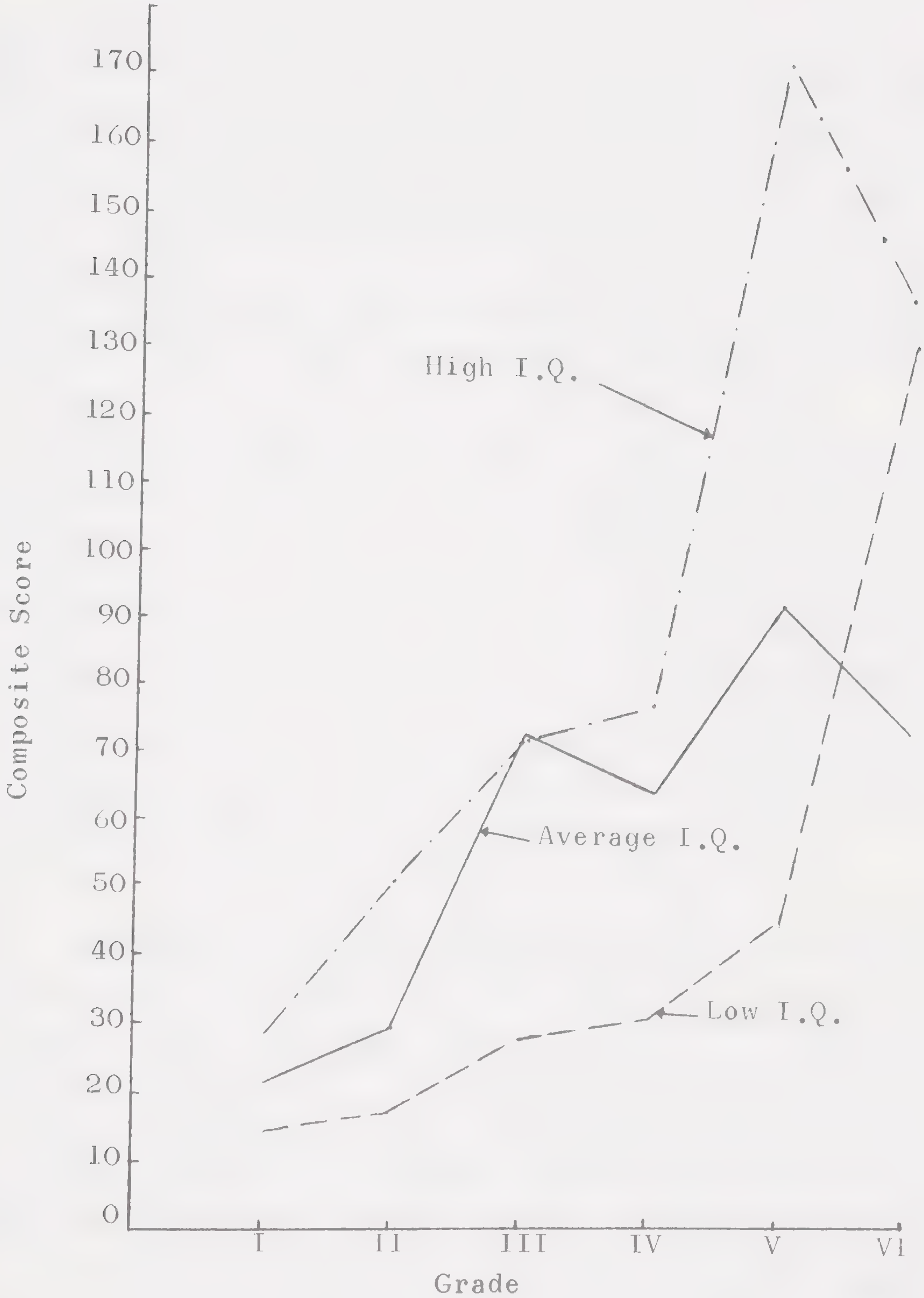

FIGURE 1

INTERACTION BETWEEN GRADE AND I.Q.
ON COMPOSITE SCORE CRITERION

page 54. The test for additivity on the power score variable yielded an F value of 0.591 which results in a level of significance of 0.707. Therefore, there is no significant interaction between sex and grade on the power score variable and hypothesis two, with respect to this variable, was not rejected. Table VI, page 54, summarizes this information.

Classification speed score. The mean speed scores for boys and girls in each grade are presented in Table V, page 54. The test for additivity on the speed score variable resulted in an F value of 0.342. This F value yields a probability of 0.887, which is far from being significant at the 0.05 level. Thus, hypothesis number two, with respect to the speed score variable, was not rejected. A summary of this information is presented in Table VI, page 54.

Composite classification score. The mean composite scores for boys and girls in each grade are presented in Table V, page 54. The test for additivity on the composite score variable yielded an F value of 0.466 which corresponds to a probability of 0.801. Therefore, there is no significant interaction between grade and sex on the composite score variable, and hypothesis two, with respect to this variable, was not rejected. This information is also given in Table VI, page 54.

Conclusion. Hypothesis number two, dealing with interaction between grade and sex, could not be rejected for

TABLE V

MEAN CLASSIFICATION SCORES FOR
BOYS AND GIRLS PER GRADE

Variable	Sex	Scores in Grade 1	2	3	4	5	6
Power scores	Boys	30.65	38.71	65.38	66.92	92.31	95.50
	Girls	28.92	49.31	60.53	66.18	75.57	89.43
Speed scores	Boys	1.51	1.89	1.40	1.30	1.04	1.16
	Girls	1.88	2.15	1.28	1.29	1.23	1.21
Composite scores	Boys	25.00	29.00	71.15	66.31	113.19	110.38
	Girls	17.77	33.69	58.71	60.18	84.00	85.69

TABLE VI

INTERACTION BETWEEN GRADE AND SEX
ON CLASSIFICATION SCORES

Variable	Source of Variance	S.S.	D.F.	M.S.	F	Prob.
Power scores	Interaction	2932.50	5	586.500	0.591	0.707
	Error	166622	168	991.799		
Speed scores	Interaction	1.253	5	0.251	0.342	0.887
	Error	123.066	168	0.733		
Composite scores	Interaction	5904.69	5	1180.94	0.466	0.801
	Error	425702	168	2533.94		

any of the three variables. Thus it was concluded that no significant interaction exists between boys and girls in grades one to six, on the criterion variables measured in this study.

Hypothesis #3

There is no significant difference between boys and girls in:

(a) Classification power score,

(b) Classification speed score,

(c) Composite classification score,

in each of the grades one through six.

The purpose of hypothesis number three was to determine whether boys or girls have greater classification abilities with respect to each of the three criterion variables within each grade. Investigation of this hypothesis was pursued by means of a two-tailed "t" test carried out by the ANOV10 IBM 360/67 computer program. A complete list of the mean scores for the boys and girls in each grade is given in Table V, page 54, and a complete résumé of the "t" values and probability levels calculated in making the comparison between boys and girls is given in Table VII, page 56.

Classification power score. None of the "t" values calculated for each of the grades one to six on the power score variable yielded a probability which was significant at the 0.05 level or better. Therefore, hypothesis number three was not rejected for any of the grades on the power score variable.

TABLE VII

COMPARISON OF BOYS AND GIRLS
CLASSIFICATION SCORES
WITHIN EACH GRADE

Grade	Variable	S.Dev. Boys	S. Dev. Girls	D.F.	T	Probability
I	Power	14.02	9.06	28	0.373	0.712
	Speed	0.82	0.79	28	-1.212	0.236
	Composite	15.40	8.72	28	1.465	0.154
II	Power	19.48	33.37	28	-1.008	0.322
	Speed	1.13	1.58	28	-0.503	0.619
	Composite	19.15	24.20	28	0.563	0.578
III	Power	36.35	24.52	28	0.421	0.677
	Speed	0.91	0.63	28	0.425	0.674
	Composite	60.93	36.44	28	0.672	0.507
IV	Power	25.28	27.38	28	0.074	0.942
	Speed	0.51	0.42	28	0.033	0.974
	Composite	47.98	38.75	28	0.374	0.711
V	Power	45.53	32.34	28	1.107	0.278
	Speed	0.53	0.75	28	-0.760	0.454
	Composite	91.60	53.84	28	1.009	0.321
VI	Power	46.09	25.13	28	0.424	0.675
	Speed	0.68	0.41	28	-0.208	0.837
	Composite	72.07	39.60	28	1.103	0.279

Classification speed score. The "t" values calculated for each grade on the speed score variable all yielded probabilities that were not significant at the 0.05 level. Therefore, _hypothesis three was not rejected for any of the grades on the speed score variable_.

Composite classification score. All the probabilities produced by the composite score "t" values were also not significant at the 0.05 level. Here too, _hypothesis number three was not rejected for any grade_.

Conclusion. Hypothesis three could not be rejected for any of the three variables in each of the grades one to six. Therefore, it must be concluded that there is no significant difference between boys and girls in their power scores, speed scores, or composite scores in any of the grades one to six.

Hypothesis #4

There is no significant difference between low, average, and high I.Q. students in:

(a) Classification power score,

(b) Classification speed score,

(c) Composite classification score,

in each of the grades one through six.

The purpose of this hypothesis was to determine if students in different I.Q. groups within each grade, do indeed have significant differences in their classification abilities, as measured by the methods of this study. This hypothesis was

tested by a Scheffé multiple comparison of means which was calculated by use of the ANOV15 IBM 360/67 computer program. As explained in Chapter 3, pages 46-47, the significance level of 0.10 is sufficient for acceptance or rejection of an hypothesis on the basis of the Scheffé test. Table III, page 50, provides the mean classification scores for each I.Q. group in every grade, and a summary of the probability levels on the Scheffé multiple comparison of means is to be found in Table VIII, page 59.

Classification power score. Significant differences in this variable were found between the high I.Q. group and the average and low groups in both grade one and grade two. The difference between the high and low group in grade one was very significant ($p<0.01$). The differences between the average and low groups in these grades, however, were not significant. On the contrary, in grade three the high and average groups were both significantly different from the low group (at the 0.01 level for the high-low comparison) but were not significantly different from each other. No significant differences were found between any of the groups with respect to this variable in the remaining upper elementary grades. See Table VIII, page 59, for a complete presentation of the probability figures.

Classification speed score. There were no significant differences in the classification speed scores of the I.Q. groups in any grade other than grade five. In grade five the

TABLE VIII

SCHEFFÉ MULTIPLE COMPARISON OF MEANS
OF I.Q. GROUPS PER GRADE

Variable	I.Q. Groups Compared	Probability level in grades 1	2	3	4	5	6
Power score	High-Average	0.031	0.017	0.289	0.999	0.382	0.155
	High-Low	0.002	0.019	0.010	0.349	0.166	0.992
	Average-Low	0.115	0.819	0.065	0.228	0.619	0.202
Speed score	High-Average	0.982	0.612	0.387	0.995	0.725	0.259
	High-Low	0.819	0.464	0.945	0.550	0.025	0.967
	Average-Low	0.843	0.866	0.617	0.364	0.035	0.404
Composite score	High-Average	0.603	0.097	1.000	0.853	0.067	0.067
	High-Low	0.211	0.027	0.290	0.187	0.013	0.967
	Average-Low	0.491	0.459	0.162	0.235	0.356	0.123

Underlining indicates significance at the 0.10 level or better

low I.Q. group had a mean speed score significantly higher ($p<0.05$) than either of the average or high groups. The average and high groups were not significantly different from each other. See Table VIII, page 59, for a complete presentation of the probability levels.

Composite classification score. Significantly different composite scores were discovered in grades two, five, and six. In both grades two and five the high I.Q. groups had significantly higher scores than either of the average or low groups. There were no significant differences between the average and low I.Q. groups. In grade six the only significant difference which existed was between the high and average I.Q. groups, the low group having obtained a higher score than the average group. See Table VIII, page 59, for a complete presentation of the probability levels.

Conclusion. The significant differences arising in speed scores and composite scores between the I.Q. groups seem hardly consistent enough to warrant rejection of hypothesis four, except in isolated cases. The most promising pattern of significant differences arises in the power score variable. Hypothesis four may be rejected with respect to this variable in each of the grades one, two, and three. The mean scores for the I.Q. groups with respect to this variable in grades four, five, and six were even farther apart than in the primary grades. However, the variation in scores rose so sharply in these grades that the large mean differences

were not statistically significant. Figure 1, page 52, presents a graphic illustration of this situation.

Hypothesis #5

There is no significant difference between students in:

(a) Classification power score,

(b) Classification speed score,

(c) Composite classification score,

in each of the grades one through six.

The purpose of this hypothesis was to determine if there is, in fact, a continual improvement in the classification abilities of students (as measured by the methods of this study) from grade one to grade six. This hypothesis was also tested by the Scheffé multiple comparison of means as calculated by use of the ANOV15 IBM 360/67 computer program. The 0.10 level of significance was used as a basis for the acceptance or rejection of this hypothesis. Table IX, page 62, includes the mean, variance and standard deviation of each of the classification variables per grade. Table X, page 63, gives the Scheffé probability matrices for the multiple comparison of classification score means for grades one to six.

Classification power score. In no case was the mean power score of any grade found to be significantly different from that of any adjacent grade. However, every mean grade score was significantly different from at least two of the other non-adjacent grade scores. The grade one mean score

TABLE IX

MEAN, VARIANCE, AND STANDARD DEVIATION
OF CLASSIFICATION SCORES PER GRADE

Variable	Grade	N	Mean	Variance	S.Dev
	1	30	29.90	152.78	12.36
	2	30	44.37	826.59	28.75
Power score	3	30	62.63	950.66	30.83
	4	30	66.50	726.05	26.95
	5	30	84.50	1720.53	41.48
	6	30	92.67	2486.44	38.55
	Total	180	63.43	1410.68	37.56
	1	30	1.67	0.71	0.84
	2	30	2.03	2.01	1.42
Speed score	3	30	1.33	0.61	0.78
	4	30	1.30	0.22	0.47
	5	30	1.13	0.44	0.66
	6	30	1.18	0.33	0.58
	Total	180	1.44	0.79	0.89
	1	30	21.87	186.47	13.66
	2	30	31.50	505.71	22.49
Composite score	3	30	64.10	2481.89	49.82
	4	30	62.83	1921.73	43.84
	5	30	99.57	6248.39	79.05
	6	30	98.83	3779.87	61.48
	Total	180	63.12	3321.07	57.63

TABLE X

SCHEFFÉ PROBABILITY MATRICES FOR EACH OF
THE CLASSIFICATION SCORES PER GRADE

Variables	Grade	1	2	3	4	5	6
Power score	1	1.000	0.668	0.007	0.002	0.001	0.001
	2	0.668	1.000	0.405	0.191	0.001	0.001
	3	0.007	0.405	1.000	0.999	0.203	0.020
	4	0.002	0.191	0.999	1.000	0.423	0.067
	5	0.001	0.001	0.203	0.423	1.000	0.960
	6	0.001	0.001	0.020	0.067	0.960	1.000
Speed score	1	1.000	0.760	0.782	0.706	0.296	0.418
	2	0.760	1.000	0.077	0.054	0.006	0.013
	3	0.782	0.077	1.000	1.000	0.974	0.994
	4	0.706	0.054	1.000	1.000	0.989	0.998
	5	0.296	0.006	0.974	0.989	1.000	1.000
	6	0.418	0.013	0.994	0.998	1.000	1.000
Composite score	1	1.000	0.990	0.065	0.081	0.001	0.001
	2	0.990	1.000	0.281	0.327	0.001	0.001
	3	0.065	0.281	1.000	1.000	0.193	0.214
	4	0.081	0.327	1.000	1.000	0.161	0.179
	5	0.001	0.001	0.193	0.161	1.000	1.000
	6	0.001	0.001	0.214	0.179	1.000	1.000

Underlining indicates a significance at the 0.10 level

was not significantly different from the grade two mean score, but it was lower than each of the grade three, four, five, and six mean scores at the 0.01 level of significance. The grade two mean score, on the other hand, was significantly lower than the mean scores of only grades five and six. The grade three and four mean scores were both significantly different only from the grade one and six mean scores. The grade five mean score was significantly higher than both that of grade one and two, and the grade six mean score was significantly higher than each of the mean scores of grade one, two, three, and four.

It was noticed that the three least significant differences in the mean power scores occurred between grades five and six (probability 0.960), grades three and four (probability 0.999), and between grades one and two (probability 0.668). The same situation existed with respect to the scores on the other two variables as well. This seemed to indicate that a natural grouping would be achieved by: combining the scores of children in grade one with those of children in grade two to make one group, combining the children's scores in grade three and four to make a second group, and finally combining grade five and six scores to make a third group. This procedure was followed and the mean scores of each new group were again analysed by means of the Scheffé test (see Tables XI and XII, page 65). The results of this test showed that the power score means of the three new groups were each significantly different from the other two ($p<0.01$). On the basis of this information, hypothesis number five was rejected with respect

TABLE XI

MEAN, VARIANCE, AND STANDARD DEVIATION
OF CLASSIFICATION SCORES PER
DOUBLE-GRADE GROUP

Variable	Group*	N	Mean	Variance	S. Dev.
	1	60	37.13	534.59	23.12
Power	2	60	64.57	827.95	28.77
score	3	60	88.58	1593.27	39.92
	Total	180	63.43	1410.68	37.56
	1	60	1.85	1.37	1.17
Speed	2	60	1.31	0.41	0.64
score	3	60	1.16	0.38	0.61
	Total	180	1.44	0.79	0.89
	1	60	26.68	363.81	19.07
Composite	2	60	63.47	2164.90	46.53
score	3	60	99.20	4929.29	70.21
	Total	180	63.12	3321.07	57.63

*Group 1 = grades one and two, group 2 = grades three and four, group 3 = grades five and six.

TABLE XII

SCHEFFÉ PROBABILITY MATRACES FOR EACH
OF THE CLASSIFICATION SCORES
PER DOUBLE-GRADE GROUP

Variable	Comparison of groups 1 & 2	1 & 3	2 & 3
Power score	0.001	0.001	0.001
Speed score	0.003	0.001	0.599
Composite score	0.001	0.001	0.001

Underlining indicates a significance at the 0.01 level

to the power score variable for the double-grade groups.

Classification speed score. Far fewer significant differences existed between the grades on the speed score variable. The high mean speed score for the second grade was significantly different from those of grades three, four, five, and six, but no significant differences existed between any of the other grades with respect to this variable (see Table X, page 63). However, when grades one and two, three and four, and five and six were grouped together and analysed by means of the Scheffé test, the grade one-two mean proved to be significantly larger than both the grade three-four mean and the grade five-six mean. The difference between the grade three-four mean and the grade five-six mean was not statistically significant. The means and probabilities discussed here are to be found in Tables XI and XII, page 65. On the basis of the preceding results, hypothesis five was rejected for the double-grade groups with respect to the speed score variable.

Composite classification score. The results of the composite score variable were similar to those of the power score variable, but with slightly fewer significant differences. The grade one mean composite score was significantly lower than each of the mean composite scores of grades three, four, five, and six. The grade two score was significantly lower than that of grade five and six. The grade three and four scores were only significantly different from the grade one score. Both the grade five and six composite scores were

significantly higher than those of grades one and two.

When the grades were broken into double-grade groups, the mean score of each group was found to be very significantly different from those of the other two groups ($p<0.01$). Tables XI and XII, page 65, summarize this information. On the basis of these results, hypothesis number five was rejected for the double-grade groups with respect to this variable.

Conclusion. Although significant differences were not found between each of the grades on each of the three variables, enough significant differences did exist to indicate a pattern of growth with respect to the power and composite variables. This growth pattern was found to be very significant when the children were grouped on the basis of double-grade groupings. Hypothesis five may be rejected with respect to all variables on the basis of this grouping technique.

It was noted, in the comparison of single grades, that a "leveling-off" took place between grade three and four on each of the three variables. This situation is shown graphically in Figure 2, page 68 (for discussion of the significance of this phenomenon see page 78).

DESCRIPTION OF THE CRITERIA USED

The classifications performed by the children in the sample for this study made use of twenty-nine different criteria. The three most popular criteria overall were those of color, liquid vs. solid state, and grain size. In total number

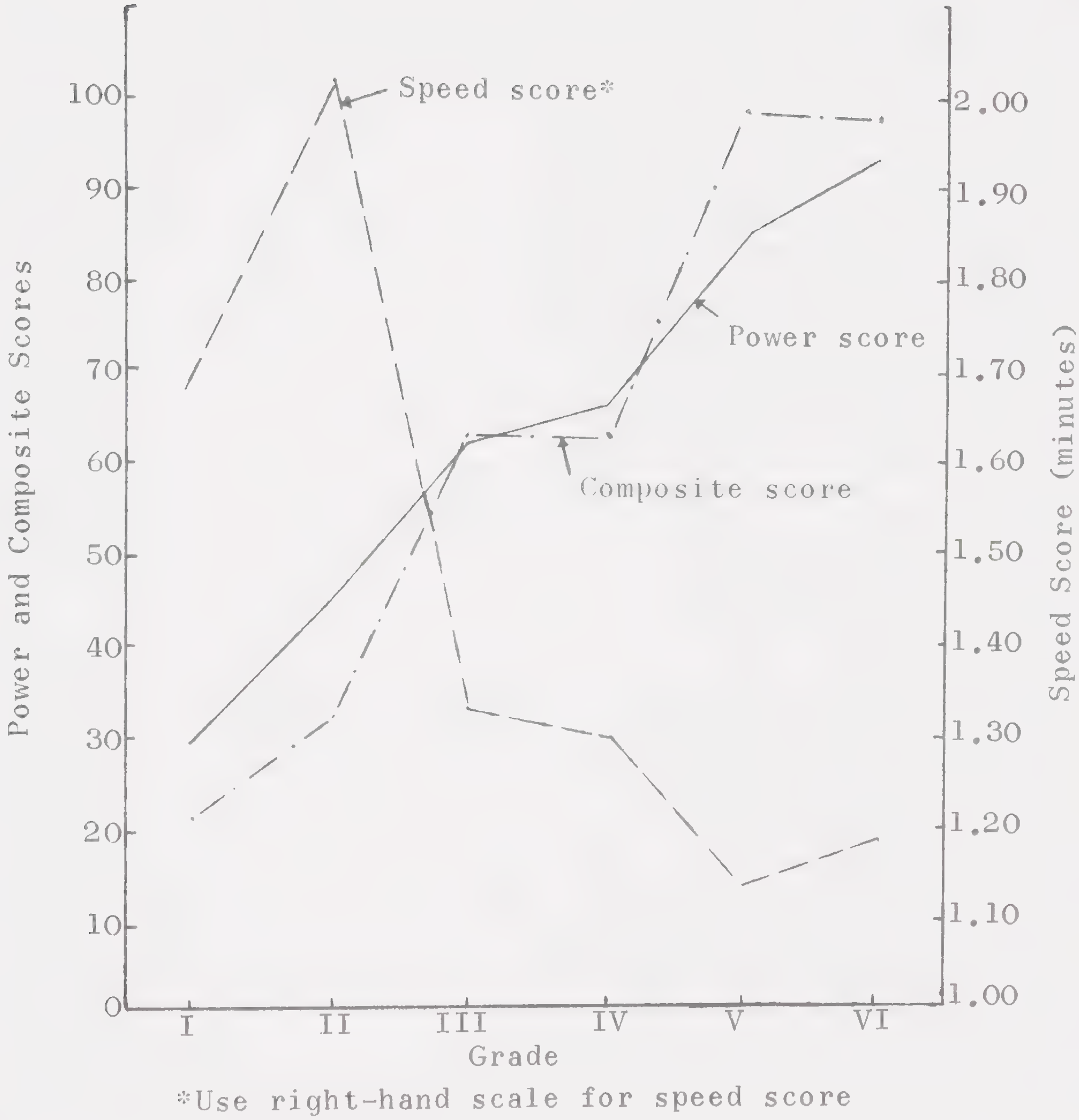

FIGURE 2

MEAN POWER, SPEED, AND COMPOSITE
SCORES PER GRADE

of classifications, color and liquid vs. solid state were very close with 165 to 163 classifications respectively. To determine true preference in this matter it was necessary to examine the order of choice for these criteria. It was then found that liquid vs. solid state was chosen first seventy-three times and second fifty-six times compared to forty-seven first and thirteen second choices for color. Thus, liquid vs. solid state seemed to be the most preferred criterion overall. Grain size remained a moderate third choice with thirty-two first choices out of a total of 147 classifications. For a complete list of the classifications used, as well as the total number of each by grade and sex, see Table XIII, page 70.

In analysing criterion preference per grade, it was found in grade one and grade two that color was a slightly more popular choice, with eleven and thirteen first choices respectively, followed closely by liquid vs. solid state with ten and eight first choices respectively. Grain size still remained a clear third. In grade three, the liquid vs. solid criterion took the lead with twenty-nine total classifications, eleven of which were first choices. Color and grain size were close to being tied for second preference with color having nine first choices and a total of twenty-four compared to grain size with seven first choices and a total of twenty-six. The criterion of weight made a large gain in total choices, growing from nine total choices in grade two to twenty in grade three. In grade four the liquid vs. solid

TABLE XIII*

NUMBER AND KINDS OF CRITERIA
USED PER GRADE AND SEX

Criteria	Grade 1	2	3	4	5	6	Totals B	G	All
Color	23	29	24	29	30	30	83	82	165
Liquid vs. Solid state	22	25	29	28	30	29	79	84	163
Grain size	17	20	26	27	29	28	69	78	147
Weight	9	9	20	20	28	21	55	52	107
Viscosity	2	8	9	13	14	25	33	38	71
Touch	4	7	12	14	13	14	28	36	64
Grain shape	6	12	10	10	9	15	30	32	62
Amount	2	3	7	6	13	16	25	22	47
Cohesiveness	4	5	10	8	7	12	24	22	46
Solubility	1	5	2	5	10	14	15	22	37
Sound	3	7	7	5	8	6	18	18	36
Odour	1	2	3	7	10	13	20	16	36
Transparency		2	9	7	6	8	15	17	32
Metals	3	5	3	5	5	10	19	12	31
Usage	2		4	4	8	9	18	9	27
Lustre	2	3	6	3	1	6	12	9	21
Taste		1		7	3	7	12	6	18
Elements or compounds			2	2	4	6	8	6	14
Where found			3	2	3	4	11	1	12
Combustible				1	6	5	9	3	12
Breakable		1	1		2	2	4	2	6
Floating dust			1		4		3	2	5
Designs in the material		1			2	1	4		4
Meniscus			1	1	1	1	2	2	4
Acid or base					2	1	1	2	3
Stains material		1		1				2	2
Familiar			1					1	1
Evaporate				1				1	1
Conducts electricity						1	1		1
Total Classifications	101	146	190	206	248	284	598	577	1175
Average per grade	3.4	4.9	6.3	6.9	8.2	9.4	6.8	6.3	

* Criteria rank ordered on the basis of the number of choices made by the <u>total population</u> for each criterion (see final column).

criterion gained in popularity, with thirteen first choices compared to six first choices for color and seven first choices for grain size. Weight still remained at twenty total choices, only one of which was a first choice. In grade five, the liquid vs. solid state and color criteria had about the same degree of popularity as in grade four. Grain size, however, dropped back to only three first choices out of a total of twenty-nine and was strongly challenged for third preference by the criterion of weight which also received three first choices out of a total of twenty-eight classifications. In grade six the liquid vs. solid state proved to be by far the most popular criterion with seventeen first choices. Grain size was next in popularity with eight first choices. Although all thirty grade sixes classified by color, only one used it for a first choice. Viscosity gained sharply in popularity from grade five to grade six, growing from fourteen total choices in grade five to twenty-five in grade six. Weight dropped back from the twenty-eight choices in grade five to twenty-one total choices in grade six, none of which were first choices. For a summary of these criterion preferences per grade see Figure 3, page 72.

Within the total sample of children there seemed to be no significant difference in criterion preferences between boys and girls. With respect to I.Q., the criterion of liquid vs. solid state was most popular in all three groups. The low I.Q. group, however, preferred color almost as much as the liquid vs. solid criterion with thirty total choices,

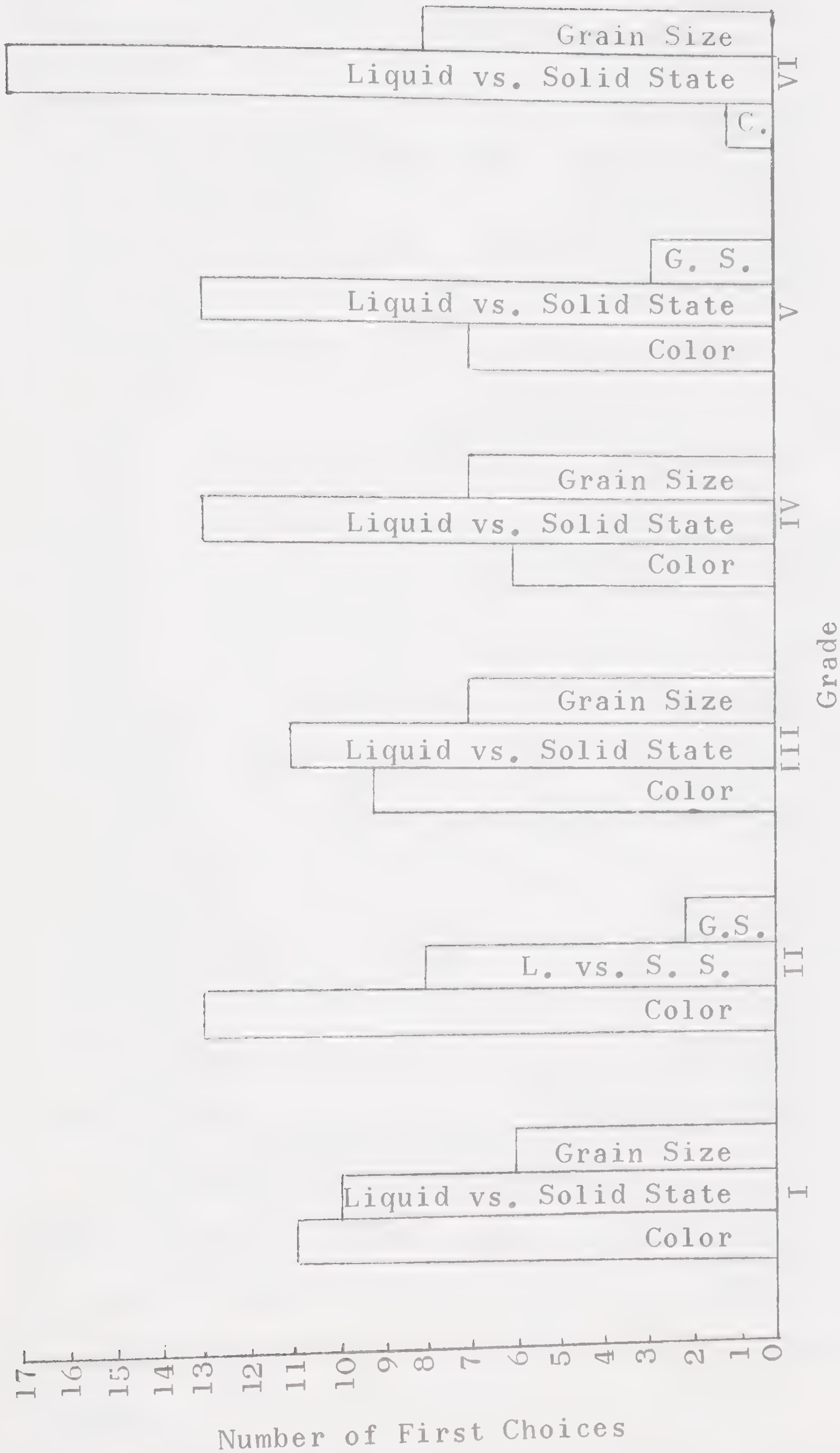

FIGURE 3

CRITERION PREFERENCE FOR EACH OF THE GRADES DETERMINED BY NUMBER OF FIRST CHOICES

twelve of which were first choices, compared to thirty-one total choices for liquid vs. solid state, thirteen of which were firsts. Color for the high I.Q. group was a remote second choice with six first choices compared with eighteen for the liquid vs. solid criterion. No other significant differences seemed to exist in the criterion preferences of the I.Q. groups.

SUMMARY OF RESULTS

Hypothesis number one examined the possibility of interaction existing between the variables of I.Q. and grade level. It was found that no significant interaction existed with respect to the power score variable and the speed score variable. Thus, hypothesis one was not rejected for these two variables. There was a significant interaction, however, on the composite score variable. This seemed attributable to a disproportionately high composite score achieved by the grade six, low I.Q. group. Hypothesis number two also examined the possibility of interaction and found none between sex and grade for any of the three criterion variables.

Hypothesis number three compared boys scores with those of girls in each of the grades one to six. In no case was there a significant difference between the boys and girls scores in any of the grades for any of the three criterion variables.

Hypothesis number four compared the scores obtained by each of the three I.Q. groups in each of the grades one

to six. On the speed score variable the only significant differences existed in grade five. Comparison of the composite scores resulted in significant differences being noted between I.Q. groups in grades two, five, and six. The power score variable yielded significant differences between I.Q. groups in each of the grades one, two, and three, however, the differences in grades four, five, and six were not significant.

Hypothesis number five compared the scores in each grade with those of each of the other grades. A significant growth pattern from grade one to grade six did emerge on the power and composite variables. When the first two grades as a group were compared with the second two grades and the third two grades, very significant differences were found between their means.

The children in this study made use of twenty-nine different criteria in their classifications. Color was the most popular criterion in grades one and two, but liquid vs. solid state became the most popular criterion in grades three, four, five, and six, and was the most popular overall criterion. There was no significant difference in criterion preference between boys and girls. With respect to I.Q. criterion preferences it was found that the criterion of color was a much more popular second choice for the low I.Q. group than it was for the high I.Q. group. The criterion of liquid vs. solid state, however, was still the most popular choice for all I.Q. groups.

Chapter 5

SUMMARY, CONCLUSIONS, IMPLICATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

SUMMARY

The purpose of this study was to measure the classification abilities of children in grades one to six. This information was then used to examine the development of these abilities throughout the elementary school years, as well as to compare the criterion preferences of each particular grade with each of the other grades.

This study was carried out on a sample of one hundred and eighty elementary school students, ninety each from the Edmonton Public and County of Strathcona school systems. The sample was composed of thirty randomly selected students from each of the grades one to six.

The I.Q. level for each child was assigned on the basis of I.Q. scores recorded on his cumulative record or on the recommendation of his teacher if such records did not contain the required information. Each student in the sample was tested individually by the investigator in a private room in the school the child attended. Each student was presented with a set of twelve small bottles which he was asked to classify in as many ways as possible on the basis of the

substances contained in them. The child was evaluated on the basis of the total number of classifications he made and on the average amount of time taken to complete his first five classifications (see page 41 for explanation). A score relating these two factors (the composite score) was then calculated.

The data thus obtained were subjected to an analysis of variance calculated by means of the ANOV10, 15, and 25 IBM 360/67 computer programs provided by the Division of Educational Research Services of the University of Alberta.

CONCLUSIONS

The specific conclusions relating to each hypothesis are found in Chapter 4. On the basis of this information the following general conclusions were formulated:

The classification power score variable seemed to differentiate both the I.Q. groups and the grade levels better than the other two variables. With respect to I.Q., the high I.Q. groups seemed to score higher than the average I.Q. groups, and the average I.Q. groups better than the low I.Q. groups with respect to this variable. This was especially true in the primary grades. An increase in power score variance in the upper elementary grades precluded many of the differences there from being statistically significant. Little difference in the time factor was found to exist between the I.Q. groups. Relatively fewer significant differences were found in the composite score variable than in the power score variable. It was also concluded that sex has no significance with respect

to classification abilities.

Hypothesis number five provided some of the most significant results of the study. Comparison of each grade and of double-grade groups indicated that there is a gradual and continual significant improvement in classification ability extending at least to grade six. These results confirm Croxton's (1936) findings even with regard to the increase in the range of scores in the higher grades. General agreement with Thompson (1941) also exists, although slightly higher average numbers of classifications for each group were found in this study. However, this discrepancy could well be attributed to the difference in the classification tasks presented.

A curious finding, when comparison of grades was made, was the plateau which occurred between grades three and four on all three of the classification scores (see Figure 2, page 68). This finding tends to corroborate those of Piaget (1964), who concluded that beyond the age of seven there is no significant improvement in classification ability up to and including the age of nine (which was the age limit for his sample). Reichard (1944) also confirmed the existence of a plateau around the age of nine years. However, he went on to test children up to fourteen years of age and found no further increase in the number of classifications made. The results of the present study do not confirm these latter fingings, as a significant increase in the classification power score was found to take place between the fourth and the sixth grade.

The plateau found by Piaget, Reichard, and this study may well represent a transition period in the shift from the non-graphic stage to the stage of true classification as described by Piaget (1964). This shift would logically involve a lag in the improvement of classification abilities while the child learned to anticipate what previously he could only accomplish through trial-and-error.

Since the objects which were required to be classified in this study differed markedly from those used in usual color-form preference tests, it was difficult to draw a comparison between the results of the two forms of test. However, color did have a significance in this task similar to its significance in other color-form preference tests, so a comparison of results was attempted on this basis. The property closest to "form" existing in the classification task associated with this study was that of grain shape. This criterion, however, did not enjoy any degree of popularity in any grade. The criterion that did seem to take the place of "form" in this classification task was that of liquid vs. solid state. It was found that grade one and two children most preferred color as a criterion while children above grade two most preferred the criterion of liquid vs. solid state. These results are similar to those of Thompson (1941) who found that her young group of children (grades one, two, and three) preferred color while the older group found other criteria just as easy to use as color. The results of this study, however, do not confirm the findings of Corah (1964)

or Suchman and Trabasso (1966), who concluded that children at least six years of age prefer form to color. The results of this study, unlike those of Kagan and Lemkin (1961), found sex to be of no significance in the criterion preference of children in the sample involved, nor was there any evidence of "meaning dominance" as found by Sigel (1954).

The results of this study indicated that color was a stronger second choice criterion in the low I.Q. group than it was in the high I.Q. group, although both groups preferred liquid vs. solid state as a first choice.

It must be concluded that although there is a difference in criterion preference in different grades, there is no difference in criterion preference between the sexes and only a slight difference between the high and low I.Q. groups.

IMPLICATIONS

On the basis of this study, certain implications can be seen:

The average number of classifications arrived at per grade were determined in the course of this study. These figures may be helpful as a rough comparison for the classification abilities of other elementary classes as determined by a similar classification task. This criterion average may also prove useful as a basis for measuring growth in the classification abilities of children exposed to the new inquiry-based elementary science curriculum. The results of this study also imply that experiences designed to improve classification

skills should continue to be presented throughout the elementary grades as no peak in classification ability is reached during this interval.

Neither boys nor girls as a group need special attention in classification skill development, however, low I.Q. students, especially in the primary grades, may need special attention as they tend to lag behind the high I.Q. groups in classification abilities. While color was still a popular criterion, this investigation showed that children, even in the lower grades, are aware of a host of other physical properties which can be used as a basis for classification. Experience with objects which exhibit many of these properties may prove to be interesting to students and may help to establish a stimulating environment for the young science student.

SUGGESTIONS FOR FURTHER RESEARCH

This study showed that somewhat less of a relationship existed between intelligence and classification ability than might have been expected. Considering this as well as the limitations imposed by the nature of the I.Q. scores obtained for this study, it would seem worthwhile to undertake further, more detailed study of this relationship. This would necessarily involve administering an intelligence test solely for the purpose of the study and correlating the scores obtained with those of the classification task.

The present study served mainly to give an indication of the present average classification ability of each of the

elementary grades. The effect on classification ability of the implementation of the new elementary science curriculum, or other changes that may take place in the future, may be determined by further research of a similar type, carried out approximately four years from the date of this study. The comparison of the results of such research with the results of this study may well indicate the change in classification ability that has taken place as a result of the implementation of curricular and/or other changes and innovations.

The new elementary curriculum guide speaks of twelve process skills, classification being but one of these. To build up a significant body of knowledge concerning all of the process skills, research of a similar type must be carried out on each of them. One of the greatest problems involved in research regarding the more complex process skills is the development of adequate measuring devices. Much research and development work will have to be done before a total picture of process skills and their acquisition in the elementary school is obtained.

Once this has been accomplished, more research will be needed to develop simple and effective measuring devices which can be administered to whole classes of children at one time, and which will give a complete profile of the developmental level of each of the process skills in each student. This investigation represents just the beginning phase of the total program of work needed to meet the challenge of evaluating the process skills of science.

BIBLIOGRAPHY

AAAS Commission on Science Education. An evaluation model and its application. miscellaneous publication, 1965.

AAAS Commission on Science Education. Commentary for teachers. miscellaneous publication, 1968.

Annett, M. The classification of instances of four common class concepts by children and adults. British Journal of Psychology, 1959, 29, 223-236.

Bennett, L. M. A study of the comparison of two instructional methods, the experimental-field method and the traditional classroom method, involving science content in ecology for the seventh grade. Science Education, December, 1965, 49, 453-468.

Bolles, M. M. The basis of pertinence. Archives of Psychology, 1937, 212, 55.

Brian, C. R., & Goodenough, F. L. Relative potency of color and form perception at various ages. Journal of Experimental Psychology, 1929, 12, 197-213.

Bruner, J. S. The act of discovery. Harvard Educational Review, 1961, 31(1), 21-32.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: John Wiley & Sons, 1956.

Butts, D., & Jones, H. Inquiry training and problem solving in elementary school children. Journal of Research in Science Teaching, 1966, 4, 21-27.

Carpenter, R. A reading method and an activity method in elementary science instruction. Science Education, April, 1963, 47, 256-258.

Corah, N. L. Color and form in children's perceptual behavior. Perceptual and Motor Skills, 1964, 18, 313-316.

Croxton, W. C. Pupils' ability to generalize. School Science and Mathematics, June, 1936, 36, 627-634.

Curriculum guide for elementary school science. Department of Education, Edmonton, Alberta, September, 1969.

de Tornyay, R. M. The effect of an experimental teaching strategy on the problem solving abilities of sophomore nursing students. Dissertation Abstracts, 1968, 28, 3499-A.

Engel, A., & Baker, H. J. Detroit beginning first-grade intelligence test: Manual of directions and key. New York: World Book Co., 1937.

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1966.

Gagné, R. M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-356.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart, and Winston, 1965.

Goldstein, K., & Scheerer, M. Abstract and concrete behavior: An experimental study with special tests. Psychological Monograph, 1941, 53(2).

Hanfmann, E., & Kasanin, J. A method for the study of concept formation. Journal of Psychology, 1937, 3, 521-540.

Heisenberg, W. The physicist's conception of nature. New York: Harcourt, Brace & Co., 1958.

Honkavaara, S. An attempt to measure intelligence in terms of organic development. Journal of Psychology, 1958, 45, 37-42.

Inhelder, B., & Piaget, J. The early growth of logic in the child. London: Routledge and Kegan Paul, 1964.

Kagen, J., & Lemkin, J. Form, color, and size in children's conceptual behavior. Child Development, 1961, 32, 25-28.

Kofsky, E. A scalogram study of classificatory development. Child Development, 1966, 37, 191-203.

Lisonbee, L. Testing, what for? The Science Teacher, 1966, 33(5), 27-29.

McCormack, M. The influence of classification and seriation ability on the mathematical achievement of first graders. Unpublished master's thesis, University of Alberta, Spring, 1969.

Mokosch, E. A process approach to science teaching. Unpublished Doctoral dissertation, University of Alberta, 1969.

Munson, H. R. Evaluating the new science teaching. Elementary School Journal, December, 1967, 68, 126-130.

Piaget, J. Judgment and reasoning in the child. London: Routledge and Kegan Paul, 1928.

Reichard, S., Schneider, M., & Rapaport, D. The development of concept formation in children. American Journal of Orthopsychiatry, 1944, 14, 156-161.

Reiner, W. B. Meeting the challenge of recent developments in science in assessing the objectives of modern instruction. School Science and Mathematics, April, 1966, 66, 335-345.

Scandura, J. The teaching-learning process: An exploratory investigation of exposition and discovery modes of problem solving instruction. Dissertation Abstracts, February, 1963, 23, 2798.

Schactel, E. Metamorphosis. New York: Basic Books, 1959.

Sigel, I. E. The dominance of meaning. Journal of Genetic Psychology, 1954, 85, 201-207.

Slinn, P. E. Teacher influence and pupil achievement. Unpublished master's thesis, University of Alberta, Fall, 1969.

Suchman, J. R. Inquiry training in the elementary school. The Science Teacher, November, 1960, 27, 42-49.

Suchman, R. G. Color-form preference, discriminative accuracy, and learning of deaf and hearing children. Child Development, 1966, 37, 439-451.

Suchman, R. G., & Trabasso, I. Color and form preference in young children. Journal of Experimental Child Psychology, 1966, 3, 177-187.

Tannenbaum, R. The development of the test of science processes. Doctoral dissertation, University Microfilms Inc., Ann Arbor Michigan, 1969.

Thompson, J. The ability of children of different grade levels to generalize on sorting tests. Journal of Psychology, 1941, 11, 119-126.

Unusual objects for classification. Instructor, January, 1967, 76, 86-87.

Welch, W. W., & Walberg, H. J. A design for curriculum evaluation. Science Education, 1968, 52, 10-16.

APPENDIXES

APPENDIX A

Process Skills (Methods of Inquiry)
(*Curriculum Guide*, 1969, pp. 7-9)

PROCESS	DESCRIPTION OF BEHAVIOR
Basic Processes	
Observing	The desired pupil behavior is increasing competence in using not only his sense of sight but also his other senses of hearing, touch, smell and taste.
Classifying	The desired pupil behavior is increasing competence in grouping articles, objects and ideas on the basis of some observable property or properties.
Quantifying	The desired pupil behavior is increasing competence in measuring lengh, weight, area, volume, and rate of change of the physical world.
Communicating	The desired pupil behavior is increasing competence in describing an experiment so that an individual who has not seen it can carry it out.
Inferring	The desired pupil behavior is increasing competence in drawing more than one inference from a set of data, demonstrating that inference can be tested by further observation, and demonstrating that an inference can be tested by applying known tests to the properties of objects. Pupils should indicate that they are able to distinguish between observations and inferences.
Predicting	The desired pupil behavior is increasing competence in conducting experiments to test predictions of relationships between two measurable quantities.

Integrated Processes

Formulating Hypotheses	The desired pupil behavior involves developing increasing competence in stating a hypothesis regarding causes of a phenomenon or the relationship between two variables. A hypothesis tells how to observe an expected outcome of an experiment.
Making Operational Definitions	The pupil should demonstrate increasing competence in stating the minimum things to do or look for in order to identify the subject being defined.
Controlling and Manipulating Variables	The desired pupil behavior is increasing competence in arranging conditions so as to be able to deliberately control and manipulate objects or conditions in an experiment.
Interpreting Data	The desired pupil behavior is increasing competence in getting the most out of data without over-simplifying, drawing conclusions supported by the data, and considering alternative explanations.
Formulating Models	The desired pupil behavior is increasing competence in building both physical and mental models to account for phenomena.
Experimenting	The desired pupil behavior is increasing competence in planning, conduction and communicating experiments in which the problem is clarified, hypotheses are stated, observations are made, and data is interpreted. This process depends upon the pupil being able to use all of the other processes.

APPENDIX B

Data Sheet

I.D.# ______________ Name ______________________________________

School ____

Grade ______________ Sex ________________ I.Q. ___________________

Power Score _______ Speed Score _______ Composite Score _______

Criteria	Order	Time	Mis.	Groupings
color				
grain shape				
grain size				
weight				
liquid vs. solid state				
metals				
viscosity				
odour				
taste				
touch				